

# EXHIBIT 4

**UNITED STATES DISTRICT COURT  
NORTHERN DISTRICT OF CALIFORNIA**

-----  
**CHARLES RIDGEWAY, et al.**

**Plaintiffs,**

**v.**

**WAL-MART STORES, INC.**

**Defendant**  
-----

**Case No. c 08-05221 SI**

F.R.C.P. RULE 26(a)(2)(B) REPORT OF JONATHAN WALKER

## Contents

I.	INTRODUCTION, QUALIFICATIONS AND ASSIGNMENT .....	4
II.	SUMMARY OF OPINIONS.....	7
III.	FACTUAL BACKGROUND.....	15
A.	WAL-MART’S PAY SYSTEM FOR PRIVATE FLEET DRIVERS .....	15
B.	PLAINTIFFS’ ALLEGATIONS AND REMEDIES SOUGHT .....	16
C.	HETEROGENEITY OF CLASS MEMBERS’ EXPERIENCES.....	17
D.	REST BREAKS.....	30
E.	CONCURRENT ACTIVITIES.....	31
IV.	PLAINTIFFS’ SURVEY AND SURVEY RESULTS.....	33
A.	PLAINTIFFS’ SURVEY INSTRUMENT .....	33
B.	HETEROGENEITY OF REPORTED EXPERIENCES.....	42
C.	SAMPLING ERROR.....	48
V.	THE UNRELIABILITY OF RECALL AS A MEASURE OF THE FREQUENCY AND DURATION OF REPEATED ACTIVITIES OR BEHAVIORS .....	51
A.	RECALL BIAS.....	52
B.	RECALL WINDOW .....	56
C.	DISTORTED PERCEPTIONS OF TIME.....	59
VI.	FURTHER FLAWS AND ERRORS IN PLAINTIFFS’ SURVEY.....	60
A.	NONRESPONSE BIAS .....	60
B.	SOCIAL DESIRABILITY BIAS.....	65
C.	DEMAND ARTIFACTS .....	65
VII.	SUMMARY OF FLAWS RELATED TO PLAINTIFFS’ SURVEY DATA.....	66
VIII.	DR. PHILLIPS’ DAMAGES AND PENALTY ANALYSES .....	69
A.	LAYOVERS .....	69
B.	PRE-TRIP INSPECTIONS.....	72
C.	POST-TRIP INSPECTIONS.....	74
D.	PAID REST BREAKS.....	76
E.	REFUELING AT WAL-MART LOCATIONS .....	79
F.	OTHER TRIP BASED COMPUTATIONS .....	82
G.	SMALL SAMPLES – DISPARATE EXPERIENCES.....	83
H.	AGGREGATE DAMAGES .....	85

I. PENALTIES..... 86

IX. CONCLUSION ..... 90

## **I. INTRODUCTION, QUALIFICATIONS AND ASSIGNMENT**

1. I am an economist. Labor economics is one of my areas of expertise. I have a bachelor's degree in economics from the University of California at Berkeley and a doctorate in economics from the Massachusetts Institute of Technology. As part of my undergraduate and graduate training, I took advanced courses in statistics and econometrics, the use of statistical tools to measure economic phenomena.

2. I am currently the President and Chief Executive Officer of Economists Incorporated ("EI"). EI was founded in 1981 for the primary purpose of providing microeconomics-related advice to individuals, corporations, non-profit organizations and governments. Among other business activities, EI regularly advises law firms and litigants about economics issues that arise in the context of litigation.

3. In addition to managing EI, I personally consult about proof of injury, quantification of damages and other economics-related topics in wage and hour cases as well as other types of litigation matters. I have testified at trial and deposition regarding proof of injury and quantification of damages wage-related damages. I have also testified at trial regarding the validity of statistical inferences drawn from survey data.

4. Gibson Dunn & Crutcher has retained me on behalf of Wal-Mart Stores, Inc., ("Wal-Mart") to conduct analyses related to injury and damages in this matter and to testify about my conclusions. Gibson Dunn & Crutcher has asked me to consider generally whether fact of injury and quantification of damages can be reliably proven and measured for the plaintiff class as a whole and for each of its members. Relatedly, Gibson Dunn & Crutcher has asked me to opine about whether the damages and penalty analyses prepared by Dr. G. Michael Phillips on behalf of the class are well-founded in economic and statistical principles and whether their conclusions are reliable.<sup>1</sup>

5. In order to form my opinions in this case, I have relied on my training and experience as an economist and material in the discovery record. A complete list of the materials that I have considered in preparing this report is attached as Exhibit B to this report. My research and

---

<sup>1</sup> When I speak of Dr. Phillips' work, my comments encompass the reports signed by Sean Chasworth and Edward Garcia which he incorporated into his report.

analysis are ongoing in that I may respond to any new analyses or assertions made by Dr. Phillips, Plaintiffs or any other experts whom Plaintiffs may retain if those analyses or assertions fall within my area of expertise. Moreover, if additional documents or information are made available to me, I will review such documents and information and that review may affect my opinions.

6. EI is being compensated for my work in this case at my standard hourly billing rate of \$625. Other economists and research staff at EI have assisted me on this matter. EI is being compensated for their time at their standard hourly rates which range from \$270 to \$535 per hour. Neither my compensation nor EI's compensation for work on this matter depends in any way on the litigation outcome.

7. The remainder of this report is organized as follows. Section II summarizes my opinions. Section III discusses the relevant factual background. That section discusses Wal-Mart's pay system and Plaintiffs' allegations about that pay system. It discusses evidence that the frequency and duration of the purportedly uncompensated activities at issue in this case, the frequency and duration of rest breaks and the frequency and duration of layover time spent in Wal-Mart trucks varied over time, by class member and by distribution-center-domicile. The section explains how and why Wal-Mart's business records are not only inadequate to prove damages in this case, but also tend to refute that reasonably accurate damages estimation in this case is susceptible to common proof. Finally, the section discusses evidence that class members engaged in activities concurrently and explains why that is relevant to proof of injury and quantification of damages.

8. Section IV discusses Plaintiffs' survey. I describe the survey instrument and the related depositions. I also discuss the survey results and what they would imply if taken at face value. I discuss proof that the survey respondents were confused by the survey and that different respondents interpreted the questions differently. I explain why this confusion and ambiguity renders the survey results unreliable for proving damages. Taken at face value, the survey indicates widely varying experiences across the class in terms of the frequency and duration of the allegedly uncompensated activities. Relatedly, I also explain that the survey averages are subject to a high margin of error if taken at face value. Taken together these two facts imply that class averages are unreliable indicators of any particular class member's experience and that the sample averages from the survey are not particularly reliable indicators of the class averages.

9. Section V discusses how and why the fundamental premise of Plaintiffs' survey and their damages estimation methodology is false, that people's recollections of the frequency and duration of past activities are a reliable and accurate measure of how frequently those activities actually occurred and how long they actually took. As Plaintiffs' underlying premise is false, Plaintiffs' survey results and class members' deposition testimony would not have been reliable for proving injury and quantifying damages even if there had been very little variation among class members in their work routines. The empirical literature establishes that people *systematically* misremember or miscalculate the frequency of past events and misperceive the passage of time. This is not random error that cancels out over sufficiently large sample sizes, but rather it is systematic, statistical bias that cannot be addressed merely by surveying more people. Sometimes sample averages overstate actual frequencies by over 100%. Average responses regarding durations may also be off by 100% or more. These two errors would compound each other in the context of this case. Consequently, leaving aside the other flaws in Plaintiffs' methodology, any damages estimate based on Plaintiffs' survey or class members' recollections could be off by an order of magnitude merely because people's recollections of the frequency and duration of past events are inaccurate and statistically biased. To my knowledge, there is no way to determine the amount of error in Plaintiffs' survey without knowing by some other means the frequency and duration of the relevant activities. This is not to say that recollections are never reliable indicators of anything, but rather that recollections about the frequency and duration of activities that were not performed at uniform frequencies and did not always take the same known and measured amount of time do not tend to be accurate.

10. Section VI identifies and explains other types of flaws and errors affecting the Plaintiffs' survey in this case. Such flaws and errors include but are not limited to confusing and ambiguous questions, nonresponse bias, social desirability bias, recall bias, sampling error and failure to address concurrent activities. These additional errors all undermine confidence in Plaintiffs' survey, and several are sufficiently severe that they are independent reasons to reject Plaintiffs' survey as unreliable. Section VII is a summary of all of the flaws I discuss regarding Plaintiffs' survey.

11. Section VIII discusses how and why Dr. Phillips' analysis is invalid and his conclusions are unreliable. The data that Dr. Phillips relies upon are inadequate to prove injury and quantify damages. Lacking reliable data, Dr. Phillips' analysis is based upon demonstrably false

assumptions, e.g., that drivers spent all 10 hours of every layover in their trucks under Wal-Mart's control, and *unreliable* data, i.e., testimony from small samples of class members, as few as two, who were not selected randomly. Section IX is a summary of my conclusions.<sup>2</sup>

## II. SUMMARY OF OPINIONS

12. In brief, there are insufficient data available in the discovery record to reliably prove injury and quantify damages on a class-wide basis with common proof. The data that are available demonstrate that there is much heterogeneity (*i.e.*, disparate experiences) among class members in terms of the frequency and duration of the allegedly uncompensated activities at issue in this case. In the class certification stage, Plaintiffs asserted that they would fill the data gap with a scientifically valid survey of the class members. The survey that they undertook was inadequate for the task. It failed to address key issues; its wording confused respondents, and it was plagued by conceptual flaws that rendered its output valueless for purposes of estimating damages. Perhaps as a consequence, Dr. Phillips abandoned it.

13. Notwithstanding the class certification declaration from Dr. Phillips' colleagues at Phillips Fractor & Co. (William Roberts and Sean Chasworth)<sup>3</sup> that Plaintiffs would rely on a statistically valid survey of class members to establish the frequency and duration of various activities when such information was not available in Wal-Mart records, Dr. Phillips makes no use whatsoever of the survey data that his firm generated in discovery. Instead, Dr. Phillips relies on arbitrary assumptions and extrapolations from selected deposition testimony from subgroups of survey-takers. The assumptions contradict the discovery record. Dr. Phillips assumed falsely that all layover time was spent in Wal-Mart's trucks under Wal-Mart's control, that no class members were ever authorized and permitted to take paid rest breaks and that none of the purportedly unpaid activities occurred concurrently with paid activities or with each other.<sup>4</sup> In fact, according to class members' sworn testimony much layover time was spent away from Wal-Mart trucks and presumably outside of Wal-Mart's control; paid rest breaks were common (based on the legal

---

<sup>2</sup> Wal-Mart disputes that it exerted control over class members during layovers even when class members spent layover time in their trucks. I have no opinion on that topic as all legal opinions are outside of my area of expertise. More generally, it is not my intention to express any legal opinions in this report, at deposition or at trial.

<sup>3</sup> Declaration of William W. Roberts, Ph.D. and Sean Chasworth In Support of Plaintiffs' Motion for Class Certification, June 27, 2014.

<sup>4</sup> I have been told by Wal-Mart's counsel to assume that a class member's "paid rest breaks" include at least the rest breaks that occurred simultaneously as Wal-Mart was paying the class member by the minute.



assumptions I was asked to make about what constitutes a paid rest break), and the purportedly unpaid activities frequently occurred concurrently with paid work and with each other. Turning to the deposition testimony, the different groups of deponents whom Dr. Phillips relied upon to estimate the frequencies and durations of rest breaks and allegedly unpaid activities do not represent random samples of the class members, so there is no statistically valid basis to extrapolate from them to the class as a whole. Unlike the statistical theory that Dr. Roberts and Mr. Chasworth testified about, theory establishing that valid inferences may be drawn from survey data that is derived from statistically random samples and used in a statistically valid way, there is no sound theory in economics or statistics validating the extrapolations that Dr. Phillips made and the inferences that he drew in this case.

14. As Dr. Phillips' analysis was based on untrue assumptions and statistically invalid extrapolations, his analysis is unreliable. Even if Dr. Phillips had accurately assessed the average frequencies and average durations of the activities at issue in this case, using those averages to estimate damages would overlook the tremendous variation across class members in the frequency at which they engaged in the activities and the amount of time they took when they did so. Most or all individual damages would be subject to significant error. There would be a high margin of error associated with aggregate damages too. I further expound upon these general opinions below.

15. ***The Wal-Mart business records produced in discovery and the survey commissioned by the Plaintiffs are inadequate to prove Plaintiffs' damages claims related to uncompensated time to any reasonable degree of certainty.*** Plaintiffs allege that class members were not compensated for certain tasks they performed. However, I am unaware that any reliable business records or other data exist showing how much time any class member (let alone all class members) spent on the specific activities that Plaintiffs say they were not compensated for. While I understand that Wal-Mart kept logs of all time worked by class members as a whole, Wal-Mart did not separately and systematically track time spent on each of the specific activities that Plaintiffs complain about. Moreover, the electronic and paper business records that have been produced in discovery suffer from various limitations that further undermine their usefulness for proving and measuring damages. Most of the types of business records exist only for a portion of the class period. None of the records measure the frequency or duration of the tasks that Plaintiffs complain about. None of the records address the problem that performance of the tasks that Plaintiffs complain about

frequently, or always, occurred concurrently with each other or with performance of tasks that Plaintiffs were compensated for.

16. ***Class members' experiences were too highly varied for the fact and quantum of damages to be susceptible to common proof.*** The business records produced in discovery indicate wildly different experiences across class members at any point in time and different experiences over time for given class members. Plaintiffs' survey and the deposition testimony of the survey respondents reinforce that class members' experiences are too varied for damages to be measured reliably on a class-wide basis.

17. ***Plaintiffs' survey does not measure the frequency at which the particular tasks at issue occurred simultaneously with other tasks, nor do the selected deposition passages that Dr. Phillips relied upon do so.*** Ultimately, Plaintiffs seek to measure time that class members allegedly worked without compensation. To do that, they need to have a reasonable estimate of the total time spent on "uncompensated" activities while class members were not concurrently engaged in other activities. For example, it is not enough to know how much time was spent on paperwork or waiting. Plaintiffs need to know how much paperwork-time was simultaneous as waiting time so that this time is not double-counted. They also need to know how much paperwork-time or waiting-time was simultaneous with compensated time such as "unscheduled time" occurring 45 minutes after commencement of a live load. Class members' survey responses do not allow the Plaintiffs to isolate the net time spent on "non-compensable" activities after deducting time spent concurrently on other allegedly non-compensable activities or other indisputably compensated activities. The selected deposition testimony that Dr. Phillips relied upon does not account fully for concurrent activities either. Most of the respondent/deponents who were asked whether their answers excluded instances in which they were simultaneously engaged in other activities said that they did not. The sole deponent who was asked the question and did not answer that way said he did not know whether he had excluded such time or not.

18. ***Neither the Wal-Mart business records produced in discovery, Plaintiffs' survey nor the selected deposition testimony that Dr. Phillips relies upon addresses how often, if at all, Wal-Mart failed to authorize or permit rest breaks.*** The business records produced in discovery do not track class members' rest breaks. Plaintiffs' survey does not ask whether drivers were authorized and permitted to take paid breaks at the rate of 10 minutes per four hours of work or major

fraction thereof, and it does not ask how frequently class members took rest breaks while simultaneously being paid for other activities. Rather, it asks about the frequency and duration of certain *unpaid* breaks. Even if the survey could measure the frequency and duration of these unpaid breaks accurately (it cannot due to the inadequacy of recall as a measure of the frequency and duration of recurrent historical events), that would tell nothing about how often Wal-Mart authorized and permitted drivers to take paid rest breaks. Class members testified at deposition that they understood that they were free to take rest breaks at their own discretion. Many testified to taking rest breaks during times when they were simultaneously receiving per-minute activity pay such as unscheduled time, or during live load or live unload. Others testified to purposefully clocking off-duty to take unpaid breaks so as to maximize the amount of their Department of Transportation on-duty time that they could apply to actually driving, and thus to maximize their pay.<sup>5</sup> The frequency and duration of unpaid rest breaks tells nothing about the frequency and duration of paid rest breaks because class members took both. The deposition testimony that Dr. Phillips relies upon does not address the frequency or duration of paid breaks either. It concerns the frequency and duration of unpaid rest breaks. Not only is this testimony unrelated to the frequency and duration of paid breaks and unrelated to whether Wal-Mart authorized and permitted paid breaks, but it is also unreliable as a measure of the frequency and duration of the deponents' own unpaid rest breaks do to recall bias let alone other class members' unpaid rest breaks.

19. ***Neither the Wal-Mart business records produced in discovery, Plaintiffs' survey nor the deposition testimony that Dr. Phillips relied upon can show whether or how often class members were underpaid during layovers.*** I understand that the Court has found that Wal-Mart should have treated class members as being engaged in compensable activity when (and if) it required them to stay with their trucks during layovers. I understand further that Wal-Mart paid class members \$42 per layover. According to Plaintiffs' randomly selected sample, class members engaged in a variety of activities away from trucks and outside of Wal-Mart's control during layovers. These activities included but were not limited to going to restaurants, visiting with friends and family, sleeping at friends' houses or at motels, attending entertainment events, exercising, shopping and resting at home. In order to calculate how much additional pay, if any, is

---

<sup>5</sup> See, e.g., Deposition of Raymond Byer 02/29/16, p. 68:18-20.

needed to bring class members' pay to at least at the minimum wage for time spent on layover at their trucks and arguably under Wal-Mart's control, it would be necessary to know how much time class members spent at their trucks during layovers. I have not seen any reliable information spanning the entire class period on this topic for any individual class member, and I am unaware of any class-wide information on this topic throughout the class period, reliable or not.

20. ***Plaintiffs' survey was ambiguous and confusing.*** For example, respondents were asked how often tasks occurred in 10 "usual trips" where a trip is defined to begin and end at a distribution center. Almost uniformly, the survey respondents reported that there was no such thing as a "usual trip." Also, "trip" is a term of art in the Wal-Mart Private Fleet Driver program, and "trips" do not consistently begin and end at distribution centers. Consequently, not all respondents interpreted the questions in the same way. Because of the poorly worded questions, respondents' answers regarding the frequency at which events occurred cannot be summarized reliably by averages or other statistics. Respondents varied in how they interpreted "usual" in light of the high variability in their experiences from day to day and over time as they moved from different distribution centers or changed the types of routes they drove. Respondents also varied in how they interpreted the word "trip" for purposes of estimating frequencies. Some respondents tried to estimate the frequency at which tasks or events occurred per 10 ten trips as defined by the survey; some tried to estimate the frequency per ten trips as trips are defined by Wal-Mart; some tried to estimate frequencies in other ways such as every two weeks, and others measured frequencies differently from question to question.

21. ***Survey respondents' answers were ambiguous and vague.*** Respondents often answered in ranges. When asked about these ranges at deposition, respondents did not uniformly report that the true answer was the midpoint of the range or even within the range, leaving unanswered what the range signifies at all. Respondents occasionally answered numerical questions with question marks or qualitative responses such as "it varies." Some respondents changed their answers during the depositions that followed immediately after filling out Plaintiffs' survey forms. Many respondents testified explicitly that their survey responses were not reliable given the variety of trips and daily experiences occurring over their tenures as Wal-Mart Private Fleet Drivers, the length of their tenures as Wal-Mart Private Fleet Drivers and the passage of time. Given these responses, it is subjective and speculative even to say what the average responses are to the various survey questions. It also demonstrates that their testimony about these very same issues

are unreliable.

22. ***Plaintiffs' survey statistics are subject to a high degree of sampling error.*** Sampling error relates to the difference between the average answer across a sample and the average answer if all members of the relevant population were polled. In their declaration in support of class certification, Dr. Roberts and Mr. Chasworth hypothesized that they could attain an estimate with a 5% margin of error from a proposed analysis of a sample of Wal-Mart business records. Plaintiffs' actual survey of class members yielded margins of error ranging as high as 72% depending on the specific question and how ambiguous responses were interpreted. Regardless of the specific question or the manner in which ambiguous responses were dealt with, no margins of error were anywhere near 5%. They were all double-digits. Sampling error is a related but still separate and distinct shortcoming from the fact that individual class members' experiences were too variable for the class average to be a good proxy for most individuals' actual experiences. Sampling error has to do with difficulty in reliably ascertaining what the class average even is.

23. ***Plaintiffs' survey results are unreliable due to recall bias, specifically people's systematic misperceptions regarding the frequency of past behaviors and the speed at which time passes.*** Plaintiffs implicitly assume that class members' recollections of the frequency at which they performed tasks and the time it took to do them are a reliable measure of the actual frequency of these tasks and time spent on them. Plaintiffs' experts provide no basis for this assumption, and the academic research tends to disprove that it is true. Surveys may be reliable means of demonstrating some types of facts, but they are not reliable means of demonstrating the particular types of facts relevant to damages in this case, namely how frequently did respondents engage in various activities and how long did those activities take. Validation studies demonstrate time and again that when people are asked these types of questions the average answer is wrong and the estimation errors are huge, sometimes several hundred percent. The fact that Plaintiffs' survey asks respondents to recall events over an eleven-plus year span further undermines its reliability.

24. ***Plaintiffs' survey is further unreliable because of nonresponse bias, social desirability bias and demand artifacts.*** Response rates to Plaintiffs' survey were low, and respondents were disproportionately former Wal-Mart employees rather than current Wal-Mart employees. Comparing answers from current Wal-Mart employees who were surveyed to answers from

former Wal-Mart employees who were surveyed, former employees tended to report engaging in the activities at issue more frequently, and they tended to report that the activities took more time. Consequently, Plaintiffs' survey appears to be afflicted by nonresponse bias that will tend to artificially inflate damages estimates. As Dr. Phillips' deponents were selected from among the survey takers, his deposition passages are affected by nonresponse too. There is also reason to be concerned that Plaintiffs' survey is biased against Wal-Mart due to social desirability bias and demand artifacts. Some of the activities at issue are activities that class members may perceive to be the right thing to do morally, particularly truck washing (a courtesy to other drivers of the same truck) and safety inspections (which reduce the likelihood of accidents). It is well established in the survey literature that survey responses tend to overstate the frequency of behaviors that are perceived to be morally right. This over reporting is referred to as social desirability bias. Finally, Plaintiffs' survey may be subject to demand artifacts. Demand artifacts refers to the tendency for people who are subjects of research to change their answers to questions or their other behavior based on their impressions of the purpose of the research. Respondents' knowledge of the purposes of the survey, their self-interest, their concern for fellow class members and their opinions about Wal-Mart may have influenced their survey responses.

25. ***Dr. Phillips' damages analysis is unreasonable and unreliable because it is also subject to recall bias, social desirability bias and demand artifacts.*** Essentially, selected class members were asked at deposition to guess about how often they performed various tasks over a period of time as long as eleven years and how long the activities typically took. Dr. Phillips extrapolates the deponents' answers to the class. As I explain above, deponents' answers are subject to incalculable error as measures of their own experiences due to recall bias, misperception of time, social desirability bias and demand artifacts. Deponents testified explicitly that they did not think that their answers were reliable due to the fact that their workdays were not uniform and that so much time had passed over the course of their careers. It would not be reliable to base damages for a single deponent/plaintiff on his admitted guess about the frequency and duration of certain of his activities over an eleven year span. Extrapolating these guesses to several hundred different class members merely exacerbates the error as there is no statistical basis to infer that the selected deponents are representative of the class, and it is implausible that they actually are. Even if the selected deponents had been a random sample, it would not be statistically reliable to extrapolate from such small samples in light of the extreme heterogeneity of experiences over time and across

the class. Assuming liability, Dr. Phillips has no statistical basis to refute that true class damages are one “ $n^{\text{th}}$ ” of his estimate where  $n$  is any integer greater than 1. In other words he has no statistical basis to rule out that class damages are 1/2, 1/10, or even 1/100 of what he estimates. On an individual basis, his analysis does not establish that any particular class members suffered any damages at all.

26. ***Dr. Phillips’ damages analysis compensates class members for time during which they were apparently relieved of all work duties.*** Although Plaintiffs dispute that layover time spent in Wal-Mart’s trucks is off-duty, to my knowledge they do not dispute that layover time that class members spent at restaurants, at hotels and motels, at home, exercising or engaging in other entertainment activities is off-duty. If time spent on such activities is not compensable, then Dr. Phillips would have had to exclude off-duty layover time to calculate layover-related damages accurately, even if one accepts the liability theory that Plaintiffs have articulated thus far. Dr. Phillips did not do so.

27. ***Dr. Phillips’ rest break analysis is apparently irrelevant and unreliable.*** Rather than measuring the shortfall in *paid* rest breaks, Dr. Phillips attempted to measure the class members’ *unpaid* rest breaks. Class members testified that they took both paid rest breaks and unpaid rest breaks, so Dr. Phillips’ calculations which focus exclusively on unpaid rest breaks cannot demonstrate that Wal-Mart failed to authorize and permit 10 minutes of paid rest during any four hour work period. Furthermore, Dr. Phillips’ rest break analysis is also unreliable because it extrapolates from selected deposition testimony of a small, nonrandom sample of class members to the class as a whole and it is subject to recall bias.

28. ***Dr. Phillips’ damages analysis is unreasonable and unreliable because it ignores concurrent activities.*** Dr. Phillips’ damages model ignores that class members were undisputedly paid while they undertook many, or all, of the activities at issue in this case. Class members also engaged in various purportedly uncompensated activities concurrently, so it would be inappropriate to sum the estimates of time spent on purportedly uncompensated activities the way that Dr. Phillips did even if the estimates were otherwise reliable. Adding them together as he does implies double-counting. By ignoring concurrent activities, Dr. Phillips overstates class damages.

29. ***Dr. Phillips’ penalty analysis is unreliable because it is based on demonstrably false***



*assumptions.* Dr. Phillips’ penalty analysis is based on the assumption that all class members suffered each of the alleged types of underpayment, except for underpaid layovers, in each pay period throughout the damages period. Dr. Phillips provides no basis for this assumption, and it is demonstrably false. Consequently, his penalty analysis is invalid. In fact, Dr. Phillips does not provide any reliable evidence that any particular class member was underpaid by any amount at all on any particular paycheck, including his or her final paycheck. Regarding waiting time penalties, Dr. Phillips makes additional demonstrably false assumptions about Wal-Mart’s electronic data causing him to overstate the number of class members in those data who terminated their employment with Wal-Mart. Regarding both waiting time and underpayment penalties, Dr. Phillips compounds his errors by extrapolating his already inflated results pertaining to class members represented in Wal-Mart’s electronic data to 90 additional class members about whom he knows essentially nothing.

### **III. FACTUAL BACKGROUND**

#### **A. WAL-MART’S PAY SYSTEM FOR PRIVATE FLEET DRIVERS**

30. Wal-Mart paid class members by activity. For most trips, driving was paid by the mile; other activities were paid on a fixed rate schedule, and still more activities were paid on an hourly basis, either as “unscheduled time” or as “scheduled time.” Certain activities are explicitly discussed in Wal-Mart pay manuals that have been produced in discovery as triggering unscheduled time, but unscheduled time was also paid for activities that were not expressly listed in these pay manuals.

31. Similarly, Wal-Mart’s pay manuals are not specific about the exact constituent components of each of the fixed rate activities, but Wal-Mart deponents and class members have testified about what tasks they understood some of the fixed-rate activities to include. Wal-Mart deponents and many class members have testified that certain activity pay is explicitly intended to compensate for many of the activities that Plaintiffs say are uncompensated. Wal-Mart witnesses and some class members say that “hook” or “depart” pay is specifically intended to compensate drivers for all of the activities directly associated with preparing for and departing a distribution center, store or vendor including pre-trip inspections, paperwork and potentially some waiting time, among other things. Wal-Mart witnesses and some class members also say that “arrive” and “drop” pay is specifically intended to compensate drivers for all of the activities directly associated with arriving at a destination and delivering a load including post-trip inspection, paperwork and potentially some



waiting time, among other things.<sup>6</sup> In addition to pay that is discussed in its manuals, Wal-Mart also delegated authority to General Transportation managers and others to augment drivers' pay on a discretionary basis through its T-Pay system if tasks took longer than expected or drivers engaged in activities without their own activity pay codes.<sup>7</sup>

## **B. PLAINTIFFS' ALLEGATIONS AND REMEDIES SOUGHT**

32. Plaintiffs allege that Wal-Mart failed to compensate class members for all of the activities that they performed, underpaid class members for layover time and failed to pay drivers minimum wage for ten full minutes of paid rest for each four hours worked or major fraction thereof. The class is defined as all persons whom Wal-Mart employed as California-based Private Fleet Drivers at any time between October 10, 2004 and the present.

33. With regard to allegedly uncompensated activities, Plaintiffs say that Wal-Mart failed to compensate class members for all time spent on pre-trip and post-trip inspections, rest breaks, wait time, fueling their tractors, washing their tractors and trailers, completing mandatory paperwork, weighing the tractors and trailers, adjusting overweight loads, meeting with driver coordinators and completing mandatory inspections. Plaintiffs seek recovery at the minimum wage rate for time spent on these activities allegedly without compensation.

34. Regarding layovers, Wal-Mart paid \$42 for layovers, an amount roughly coincident with federal per diem reimbursement for meals and incidental expenses that accrue when employees travel away from their homes for business purposes.<sup>8</sup> Plaintiffs say that layover time was compensable and that the \$42 was a wage. Layovers generally lasted at least 10 hours although not all of that time was spent with the truck. Dividing the \$42 dollar payment by 10 yields \$4.20 per hour. This is less than the California minimum wage which has ranged from \$6.75 to \$10.00 per hour over the class period. Plaintiffs seek the shortfall between the pay that class members actually received for layovers and the pay that Plaintiffs say class members should have received in order to have earned at least the minimum wage.

---

<sup>6</sup> See Wal-Mart's Opposition to Plaintiffs' Motion for Summary Judgment, March 25, 2016 pp. 17-18 for references to class member testimony that certain paperwork, inspections and wait time were directly related to and integral parts of the hook/depart and arrive/drop activities. See also Davis Dep. 163:17-164:2; Mosley Dep. 116:11-21 and 134:20-25, 174:20-176:1; and Fantasia Dep. 43:9-15.

<sup>7</sup> Mosley Dep. 106:6-24. Steven Davis Declaration, April 9, 2015, paragraph 2.

<sup>8</sup> Federal per diems varied between \$23.25 and \$71.00 for overnight stays in California during the damages period depending on the city and the year. See the U.S. General Services Administration web site at <http://www.gsa.gov/portal/category/100120>.

35. As for paid rest breaks, Plaintiffs say that it was impossible for a class member to be paid for taking an off-duty rest break under Wal-Mart's current pay plan because there is no specific fixed-rate pay code for rest breaks. (Motion for Class Certification, p. 6). As a result, Plaintiffs are seeking minimum wage for the rest breaks that were actually taken.

36. I understand that Plaintiffs also seek waiting time penalties for class members who left Wal-Mart's employ without having been paid for the allegedly uncompensated activities, without having been paid at the minimum wage rate for all layovers or without having been paid premium pay for rest breaks that were not authorized and permitted. They also seek underpayment penalties for all class members. I understand that Wal-Mart asserts various defenses to Plaintiffs' allegations, defenses which I do not necessarily discuss in this expert report.

### **C. HETEROGENEITY OF CLASS MEMBERS' EXPERIENCES**

37. There are only limited data available concerning the topics that are most important to damages estimation in this case. To my knowledge there are no electronic Wal-Mart data that span the entire class period concerning a) the frequency at which class members engaged in most of the purportedly uncompensated activities, b) the duration of such activities when they occurred, and c) the frequency at which class members took paid rest breaks. The information that is available suggests that these frequencies and durations varied substantially across class members, over time and by distribution center.

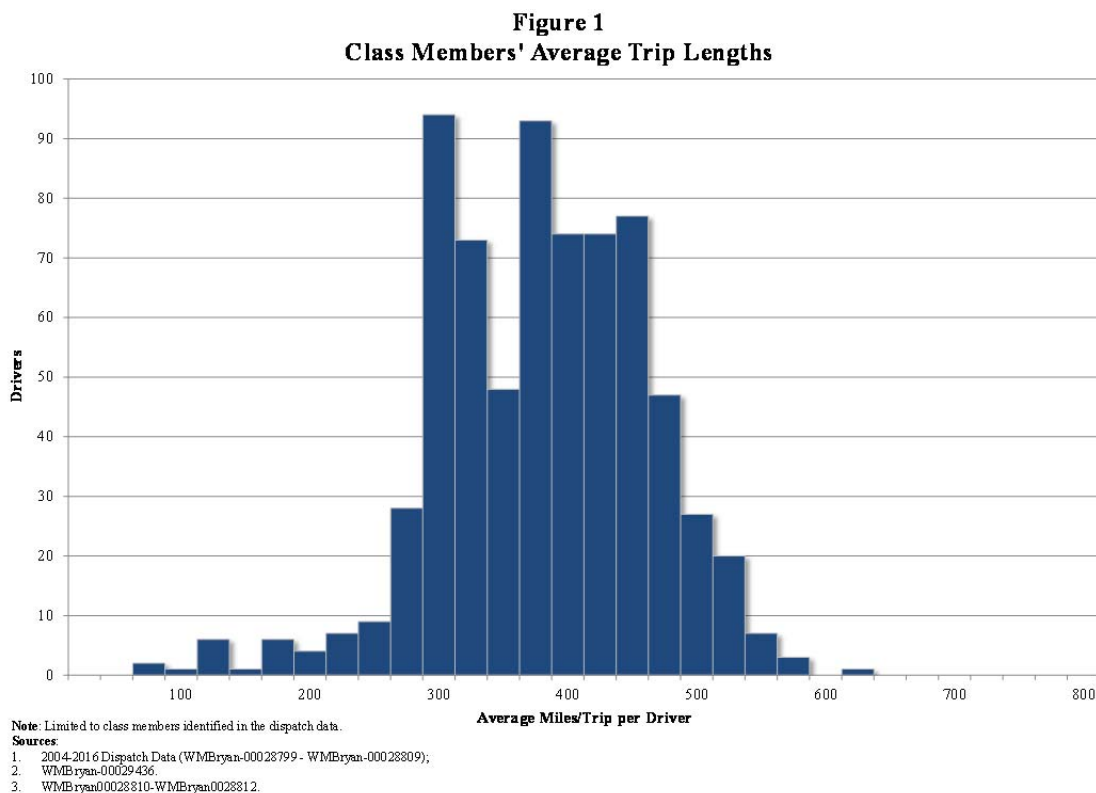
38. The class members whom Plaintiffs selected at random to complete surveys and to sit for depositions uniformly testified to experiences that varied markedly, often from day to day. The survey respondents/deponents who were asked whether there was such a thing as a usual trip or a usual week testified that there was not.<sup>9</sup> The respondents/deponents testified that their own experiences engaging in the allegedly uncompensated activities varied dramatically over time. They testified that the frequency at which they washed their trucks would depend on their routes, the time of year and the weather. They testified that waiting time would vary depending on whether the delivery or pickup location was a vendor, traditional store, grocery store or another distribution center. Given that the delivery or pickup location was a vendor, a traditional store, a grocery store or another distribution center, wait time would vary based on which vendor, which

---

<sup>9</sup> See, among others, the Depositions of Mark Alumbaugh, 03/18/16, p. 50:12-22; Frank LaCas, 02/18/16, p. 66:11-13; Greg Ryan, 02/22/16, p. 50:12-15; and Angel Vasquez, 05/04/16, pp. 67:22-68:1.

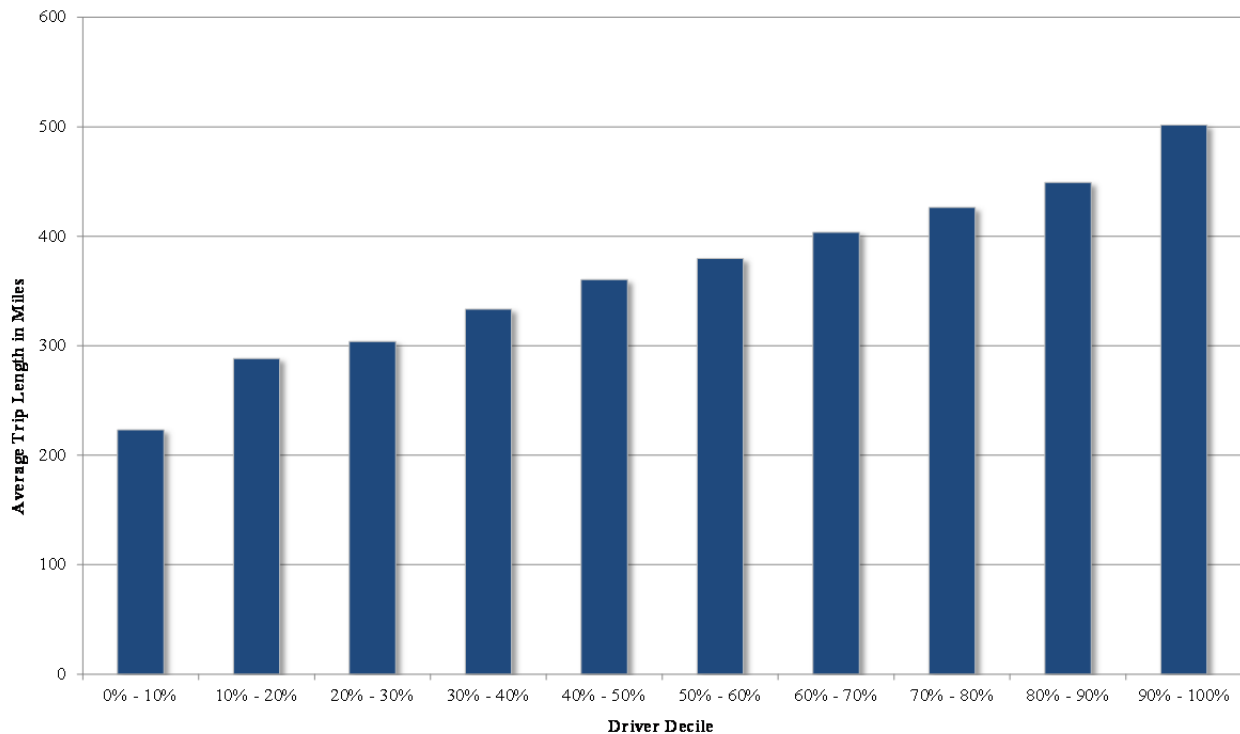
traditional store, which grocery store and which distribution center. Class members' experiences were so varied over time that the class members selected randomly by Plaintiffs were often unwilling or unable to specifically answer questions about the frequency at which they usually engaged in activities at issue in this litigation or the usual duration of those activities. I discuss this in greater detail below in the section of this report addressing Plaintiffs' survey.

39. Wal-Mart produced electronic dispatch data spanning the time period from November 2004 through the middle of January 2016. These data tend to confirm that class members' work experiences were highly variable and that it is unlikely that representative data will reliably reflect any individual's experience. Figure 1 illustrates drivers' disparate experiences in terms of average trip length. Each bar on Figure 1 represents an average trip length within a 25 mile range. The height of a bar indicates how many drivers had average trip lengths in the dispatch data within a given mileage range. If the average trip length for most drivers was within a given 25 mile range, then the bar for that range would be very tall and all of the other bars would be very short, perhaps having no height at all. To the contrary, drivers' average trip lengths varied widely. At the extremes, some drivers' average trip lengths were as low as 50 to 75 miles while other drivers' average trip length was 600 to 625 miles.



40. Although it should be clear from Figure 1 that the variance in class members' experiences is due to a general lack of similarity across all drivers and not due merely to a handful of individual drivers having very high average trip lengths and a handful having very low trip lengths, I have prepared Figure 2 to explore this topic further. Figure 2 is another chart showing how average trip length varied across drivers. Each bar in Figure 2 shows the average trip length for all trips by 10% of the class members in the dispatch data. To create Figure 2, I sorted the drivers by average trip length such that the first bar represents the 10% of drivers with the shortest average trip length; the next bar represents the 10% of drivers with the next shortest average trip lengths, *etc.* To illustrate the wide variety of average experiences across drivers, compare the bottom decile (the 10% of drivers having the shortest average trip lengths) to the top decile (the 10% of drivers having the longest average trip lengths). The bottom decile of drivers in terms of average trip length had average trips of 223 miles. The top decile of drivers in terms of average trip length had average trips of 502 miles. The difference in average trip length between the bottom decile and the top decile is 125%.

**Figure 2**  
**Class Members' Average Trip Lengths by Decile**



**Note:** Limited to class members identified in the dispatch data.

**Sources:**

1. 2004-2016 Dispatch Data (WMBryan-00028799 - WMBryan-00028809);
2. WMBryan-00029436.
3. WMBryan00028810-WMBryan0028812.

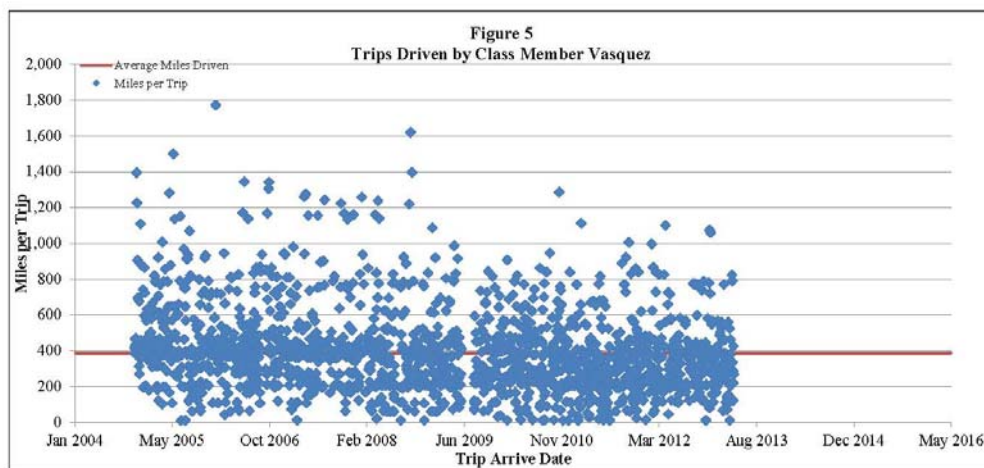
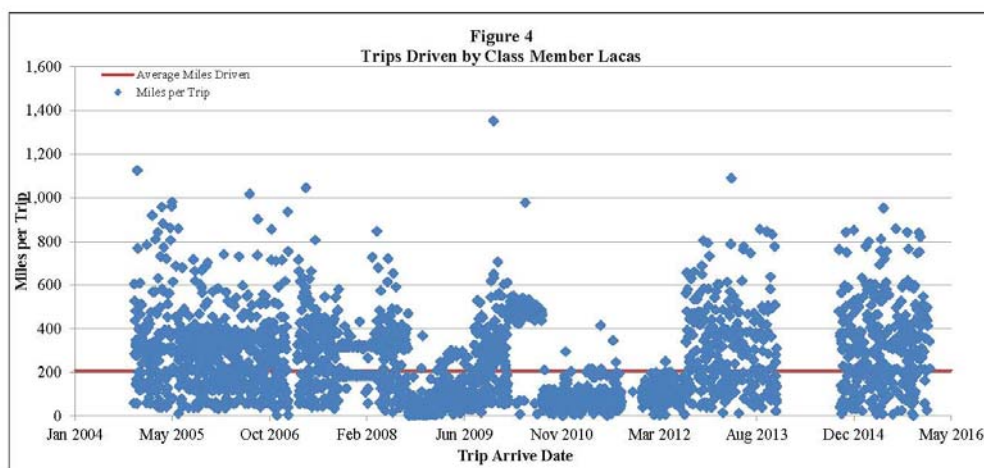
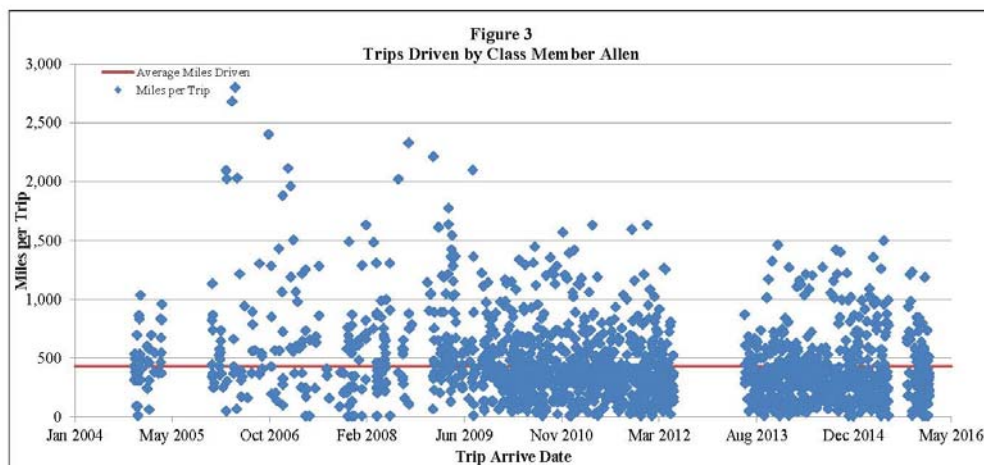
41. Figures 1 and 2 understate the variety of experiences because they are based on each driver's average trip length. Although it may be possible to calculate an average trip length for a driver, the fact that an average can be calculated does not mean that most trips are "average" or that the average is particularly informative about what typically occurs. For example, consider two people, a six foot tall adult and a newborn 20 inches long. Their average height is 46 inches or 3'10". Although the average is calculable, it is not a reasonable measure of either of their heights. Averages may blur distinctions across populations such as all trips for one driver or all trips for the class, and mask variability.

42. Figures 3, 4 and 5 illustrate how much variability there was in individual drivers' trip lengths. These figures show that for each class member, his or her average experience can mask the tremendous variety in his or her population of experiences. Figure 3 relates to the trips that Mr. Allen logged; Figure 4 relates to the trips that Mr. LaCas logged and Figure 5 relates to the trips that Mr. Vasquez logged. Mr. Allen, Mr. LaCas and Mr. Vasquez are among the class members whom Plaintiffs selected randomly for purposes of developing representative data about the class. When the 40 randomly selected class members who were surveyed and deposed are arranged alphabetically, Mr. Allen, Mr. LaCas and Mr. Vasquez are the first, the exact middle and the last survey respondents/deponents respectively.

43. Figures 3, 4 and 5 are scatter plots. The y-axis represents distance measured in miles. The x-axis represents time measured in days. Each point on Figures 3, 4 and 5 represents the length and end date of a trip for a driver, depending on the graph either Mr. Allen, Mr. LaCas or Mr. Vasquez. The red line on each of Figures 3, 4 and 5 represents the average trip length for the respective class member.

44. As a threshold matter, the average trip lengths for Mr. Allen, Mr. LaCas and Mr. Vasquez are very far apart, indicating much different experiences among the three drivers on average. The highest of the three averages, Mr. Allen's is more than double the the lowest of the three averages, Mr. Lacas'.

45. Now consider each driver individually. Figure 3 relates to Mr. Allen. If Mr. Allen's trips were the same length, then each dot on Figure 3 should occur at a height corresponding to Mr. Allen's average trip length of 431 miles. The chart should look like a horizontal line at 431, and the narrower the line, the less disparity among trips in terms of length. Figure 3 looks nothing like this. Mr. Allen had many trips that were far longer than 431 miles and many that were far shorter. The average is not "representative" of Mr. Allen's varied experiences. It is not obvious that Mr. Allen or anyone else who experienced the same trips as Mr. Allen and at the same time would be able to say accurately how long the average trip was. It seems clear

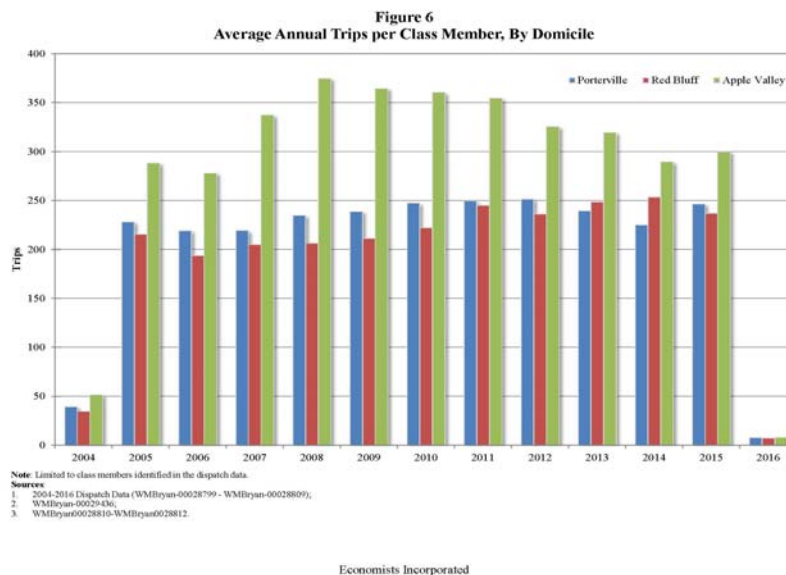
**Sources:**

1. 2004-2016 Dispatch Data (WMBryan-00028799 - WMBryan-00028809);
2. WMBryan00028810-WMBryan00028812.

Economists Incorporated

from the figure that trips varied over time so that the average over one week, month or year would not necessarily be the same as the average over any other week, month or year. Figures 4 and 5 tell similar stories about Mr. LaCas' and Mr. Vasquez's experiences as Figure 3 tells about Mr. Allen's. Trip lengths varied significantly and showed no consistent patterns over time. Most trips were far from average in terms of length, and average trip length varied over time. I have prepared similar charts for all of the survey respondents/deponents whom I have identified in the dispatch data. I have collected these in Appendix 1 to this report. They all corroborate the wide variation in trip lengths for drivers over time.<sup>10</sup>

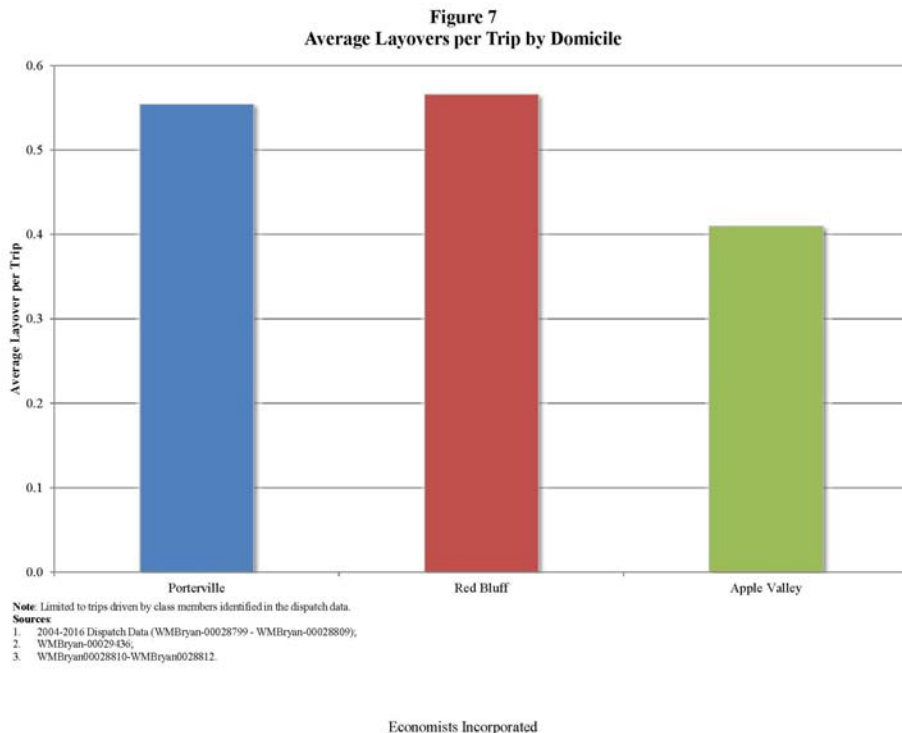
46. Wal-Mart's data also show differences among drivers domiciled at different distribution centers. Figure 6 demonstrates that average annual trips per driver vary systematically across distribution centers. Apple Valley has always had a much higher number of average annual trips per driver. For example, in 2008, Apple Valley had nearly twice as many trips per driver on average as Red Bluff. Porterville and Red Bluff have varied from year to year in terms of which domicile had higher number of average trips per year per driver.



<sup>10</sup> Mr. Allen was arguably "different" because he drove a Wal-Mart moving van from approximately November 2004 through October 2009, and this work involved some trips that were as long as 2,500 miles. The moving vans relocate Wal-Mart executives' personal belongings when they transfer to different locations. However, this difference in work experience between time assigned to the moving van and other times merely reinforces the heterogeneity of given individuals' experiences, and Mr. Allen was not alone in having dramatically different assignments during different portions of the class period. See the discussion regarding Mr. Nettles below for another example. Also see Appendix 1 showing that heterogeneity was the norm.



47. Figure 7 shows that the number of layovers per trip was much smaller in Apple Valley than in Porterville and Red Bluff. Average number of layovers per trip was 35% higher for drivers domiciled in Porterville and 38% higher for drivers domiciled in Red Bluff than it was for drivers domiciled in Apple Valley.

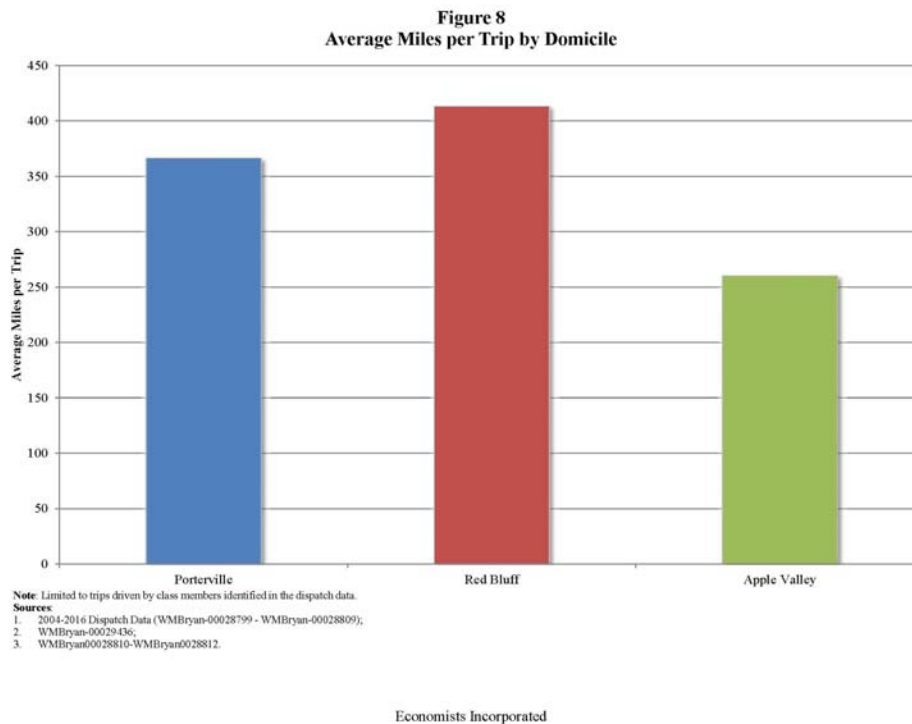


48. Figure 8 shows that average miles per trip also varied across domiciles. Over the class period, miles per trip was the lowest in Apple Valley. Average miles per trip was 41% higher in Porterville than in Apple Valley and 59% higher in Red Bluff than in Apple Valley.

49. The differences in trip lengths across class members, over time and between domiciles may affect the frequency and duration of the tasks at issue in this litigation. Differences in the number of layovers per trip is directly relevant to the aggregate amount of extra pay, if any, that Wal-Mart would owe a class member for layover time. Also, class member testimony indicates that coordinator meetings, inspection time and fueling time are also related to trip length.<sup>11</sup>

<sup>11</sup> See, e.g., Depositions of Dale McFall 05/12/16 p. 82:4-21 (coordinator meetings); Robert Garcia 03/21/16, pp. 58:22-59:19 and John Goodman 02/22/16, pp. 57:15-58:2 (inspections); Mark Alumbaugh 03/18/16, pp. 66:23-67:22 and Angel Vasquez 05/04/16, p. 78:5-14 (fueling).





50. Plaintiffs conducted a survey of class members. I discuss the survey at length below. However, it is relevant to the discussion of heterogeneity across drivers in terms of mileage that survey responses varied across drivers based on the average distance of trips driven. I divided the respondents into quartiles based on average trip length in miles and compared the average responses for each of the Plaintiffs' survey questions for the bottom quartile of drivers based on average trip length to the responses for the top quartile. I report the results in Figure 9.

51. Given a sample of 37, a quartile is only 9 respondents.<sup>12</sup> With such small sample sizes, 9 people per comparison group, only extreme differences will be statistically significant. Nevertheless, there were statistically significant differences between the two groups' answers regarding the usual length of coordinator meetings at the end of a trip, the frequency of having to wait for trucks to be loaded or unloaded and the frequency at which they reported stopping driving specifically to take unpaid rest breaks.<sup>13</sup> The low-mileage drivers reported waiting for

<sup>12</sup> One respondent never drove a trip. Two others were domiciled outside of CA for part of the class period and we have been unable to identify them in the dispatch data.

<sup>13</sup> Statistical significance and p-values related to differences in means are based on Student-t tests unless specified otherwise.

**Figure 9**  
**Statistical Tests of Survey Responses**  
**Drivers in Top Quartile vs Bottom Quartile of Average Miles per Trip**

Question	Difference of Means (A)-(B)	p-Value	Percentage Difference of Means	Top Quartile		Bottom Quartile	
				Number of Responses	Mean (A)	Number of Responses	Mean (B)
3a Usual pretrip length	-2.33	0.43	19.2%	9	12.17	9	14.50
4a Usual post-trip length	-1.22	0.59	12.3%	9	9.94	9	11.17
5a # roadside/weight inspections in usual month	2.19	0.22	399.6%	6	2.74	7	0.55
5b Usual inspection length	-4.03	0.57	21.4%	7	18.79	8	22.81
6a # washings in 10 usual trips	0.50	0.60	25.9%	7	2.43	7	1.93
6b Usual washing length	-2.00	0.83	6.7%	9	30.00	9	32.00
7a # weighings outside Wal-Mart in 10 usual trips	-0.67	0.37	53.3%	6	1.25	6	1.92
7b Usual length of weighing outside Wal-Mart	-1.39	0.85	6.3%	9	21.94	9	23.33
8a # adjustments outside Wal-Mart in 10 usual trips	-0.52	0.20	62.9%	6	0.83	7	1.36
8b Usual time to adjust	21.07	0.29	100.0%	7	42.14	7	21.07
9a # refuelings at Wal-Mart in 10 usual trips	2.60 *	0.09	67.9%	8	6.44	9	3.83
9b Usual time to refuel at Wal-Mart	0.39	0.92	2.3%	9	17.22	9	16.83
10a # refuelings outside Wal-Mart in 10 usual trips	1.31 *	0.05	167.9%	6	2.08	9	0.78
10b Usual time to refuel outside Wal-Mart	0.66	0.90	3.3%	8	20.94	9	20.28
11a # meetings at start of day or trip in 10 usual trips	1.60	0.24	24.2%	7	8.21	9	6.61
11b Usual length of meeting at start of day or trip	2.47	0.74	25.4%	8	12.19	9	9.72
12a # meetings at end of trip in 10 usual trips	0.90	0.63	16.0%	7	6.57	9	5.67
12b Usual length of meeting at end of trip	-5.04 **	0.04	109.0%	8	4.63	9	9.67
13a # times waiting for loading in 10 usual trips	-1.44 **	0.04	115.0%	6	1.25	8	2.69
13b Usual time waiting for loading	20.00	0.25	32.0%	8	82.50	6	62.50
14a # times waiting for unloading in 10 usual trips	-1.78 **	0.04	88.9%	5	2.00	9	3.78
14b Usual time waiting for unloading	15.00	0.11	39.1%	9	53.33	9	38.33
15a # rest breaks in 10 usual trips	13.60 ***	0.01	376.7%	7	17.21	9	3.61
15b Usual rest break length	0.06	0.98	0.5%	8	12.00	9	11.94

**Sources:**

- 2016 Plaintiff Questionnaires
- 2004-2016 Dispatch Data (WMBryan-00028799 - WMBryan-00028809)
- WMBryan00028810-WMBryan0028812.

**Notes:**

- The means reported are the averages of survey responses for which a mean could be computed, i.e., responses with a single number or a range of numbers.
- \* denotes significance at the 10% level, \*\* denotes significance at the 5% level, and \*\*\* denotes significance at the 1% level.
- Survey respondents who could be identified in Wal-Mart's dispatch data were divided into quartiles. The number of responses within the quartile to a given question may be less than the number of drivers in the quartile to the extent that some drivers did not respond, or provided a non-numeric response, to that question.
- If an answer is vague with a number, such as "60 depends on...", then we take the number as the answer.
- If an answer is nonnumeric such as "varies", "do not recall", "unknown", etc., we treat it as a blank.
- If an answer is in the format of ">#%" or "#+", we take # as the minimum, and blank as the maximum.
- If an answer is in the format of "<1", we take 1 as the maximum, and 0 as the minimum.
- If the answer does not match the question, we treat it as a blank.
- If an answer is "no", we assume the numeric response is 0. If an answer is "yes" and the numeric response is blank, we assume the response is blank. If an answer is "yes" and the numeric response is 0, we assume the response is 0.
- Based on deposition testimony, some drivers' survey responses were updated. Kenneth Nevarez answered "no" for some questions and then provided numeric answers. We use those numbers since he stated in his deposition that he intended to answer "yes". John Rivero answered "no" for 11 and 12, and then provided numeric answers. We take "no" as answers and treat the numbers as zeros since according to his deposition, he intended to answer "no".

unloading nearly twice as often as the high-mileage drivers. The low-mileage drivers reported waiting for loading more than twice as often as the high-mileage drivers did. The low-mileage drivers reported spending more than twice as long meeting with the coordinator at the end of a trip as did the high-mileage drivers. The high-mileage drivers reported pulling over for unpaid rest breaks over four and a half times as often low-mileage drivers.

52. The average responses to other questions were also quite different between the high-mileage and low-mileage drivers, even though the differences were statistically insignificant. High-mileage drivers reported that CHP/DOT roadside/weight station inspections occurred nearly 5 times as frequently as low-mileage drivers reported. High-mileage drivers reported that it took twice as long to make weight adjustments when making adjustments outside of a Wal-Mart facility. There were differences of 20% or more between high-mileage and low-mileage drivers to 15 out of the 24 of Plaintiffs' activity-related survey questions. I discuss in greater detail below various flaws in Plaintiffs' survey. However, this comparison of

high-mileage and low-mileage survey respondents establishes either a) that class members with dissimilar mileage experiences also had dissimilar experiences in terms of the frequency and duration of purportedly uncompensated activities, b) that Plaintiffs' survey is unreliable or c) both.

53. Similarly, comparison of answers by respondents domiciled at different distribution centers establishes either a) that Plaintiffs domiciled at different distribution centers had dissimilar experiences in terms of the frequency and duration of purportedly uncompensated activities, b) that Plaintiffs' survey is unreliable or c) both. Compare survey responses from Porterville drivers to responses from Red Bluff drivers. Again, the sample sizes are small, so only extreme differences in survey responses will be statistically significant. Red Bluff drivers reported spending 62% more time waiting for trucks to unload than Porterville drivers reported. The likelihood of a disparity this large, notwithstanding the small sample sizes, would be 3% or less if Plaintiffs' survey were valid. Porterville drivers reported that their CHP/DOT roadside/weigh station inspections took 47% longer than Red Bluff drivers reported. The likelihood of seeing such a large disparity by chance would be less than 1% if Plaintiffs' survey were valid. Porterville drivers reported having to wait for trucks to load 149% more often than Red Bluff drivers reported having to wait. The likelihood of such a large discrepancy is only 2% or less if Plaintiffs' survey is valid. Porterville drivers reported weighing at non-Wal-Mart weigh stations 118% more often than Red Bluff drivers reported. The likelihood of such a disparity is 5% or less if Plaintiffs' survey is valid. Porterville drivers reported having to make adjustments at non-Wal-Mart weigh stations 163% more often than Red Bluff drivers reported having to make such adjustments. The likelihood of such a disparity is 2% or less if Plaintiffs' survey is valid. Porterville drivers reported washing their vehicles 87% more often than Red Bluff drivers reported. The likelihood of such a disparity is 7% or less if Plaintiffs' survey is valid. Porterville drivers reported end of trip meetings with driver coordinators lasting 165% longer than Red Bluff drivers reported. The likelihood of such a disparity is 7% or less if Plaintiffs' survey is valid.

54. In Figures 10, 11 and 12, I report differences in mean survey responses to duration and frequency questions between respondents domiciled in different distribution centers. The differences between mean responses are greater than 20% for nearly half of the comparisons. Many of the differences are statistically significant despite the small sample sizes for the comparisons. Regardless of which two distribution centers are being compared to each other, there were always at least two questions for which the average answer differed by 100% or more between domiciles.

**Figure 10**  
**Statistical Tests of Survey Responses**  
**Porterville vs Red Bluff**

Question	Difference of Means (A)-(B)	p-Value	Percentage Difference of Means	Porterville		Red Bluff	
				Number of Responses	Mean (A)	Number of Responses	Mean (B)
3a Usual pretrip length	1.73	0.38	13.1%	13	15.00	15	13.27
4a Usual post-trip length	2.07	0.43	19.8%	13	12.50	15	10.43
5a # roadside/weight inspections in usual month	0.59	0.83	22.6%	6	3.19	11	2.61
5b Usual inspection length	8.55 ***	0.01	47.4%	11	26.59	13	18.04
6a # washings in 10 usual trips	1.86 *	0.07	87.2%	10	4.00	11	2.14
6b Usual washing length	3.58	0.65	13.1%	11	30.91	15	27.33
7a # weighings outside Wal-Mart in 10 usual trips	1.47 **	0.05	117.8%	9	2.72	10	1.25
7b Usual length of weighing outside Wal-Mart	0.58	0.92	2.7%	12	22.08	15	21.50
8a # adjustments outside Wal-Mart in 10 usual trips	1.44 **	0.02	162.5%	9	2.33	9	0.89
8b Usual time to adjust	-15.70	0.24	68.4%	11	22.95	13	38.65
9a # refuelings at Wal-Mart in 10 usual trips	-0.77	0.55	14.3%	13	5.38	13	6.15
9b Usual time to refuel at Wal-Mart	-2.08	0.48	13.5%	12	15.42	15	17.50
10a # refuelings outside Wal-Mart in 10 usual trips	0.30	0.76	16.9%	11	2.05	10	1.75
10b Usual time to refuel outside Wal-Mart	0.58	0.90	2.6%	13	23.08	15	22.50
11a # meetings at start of day or trip in 10 usual trips	0.35	0.77	4.2%	13	8.62	13	8.27
11b Usual length of meeting at start of day or trip	4.73	0.40	52.6%	13	13.73	13	9.00
12a # meetings at end of trip in 10 usual trips	-1.71	0.31	31.2%	13	5.46	12	7.17
12b Usual length of meeting at end of trip	7.19 *	0.07	165.5%	13	11.54	13	4.35
13a # times waiting for loading in 10 usual trips	2.60 **	0.02	148.6%	10	4.35	10	1.75
13b Usual time waiting for loading	-7.57	0.73	11.3%	10	66.75	11	74.32
14a # times waiting for unloading in 10 usual trips	0.17	0.85	7.1%	11	2.50	9	2.33
14b Usual time waiting for unloading	-20.02 **	0.03	62.0%	11	32.27	12	52.29
15a # rest breaks in 10 usual trips	-4.77	0.16	51.8%	13	9.19	12	13.96
15b Usual rest break length	2.67	0.28	23.2%	12	14.17	14	11.50

**Sources:**

1. 2016 Plaintiff Questionnaires

**Notes:**

1. The means reported are the averages of survey responses for which a mean could be computed, i.e., responses with a single number or a range of numbers.
2. \* denotes significance at the 10% level, \*\* denotes significance at the 5% level, and \*\*\* denotes significance at the 1% level.
3. The number of responses per domicile to a given question may be less than the number of survey respondents domiciled at that location to the extent that some drivers did not respond, or provided a non-numeric response, to that question.
4. If an answer is vague with a number, such as "60 depends on...", then we take the number as the answer.
5. If an answer is nonnumeric such as "varies", "do not recall", "unknown", etc., we treat it as a blank.
6. If an answer is in the format of ">#" or "#+", we take # as the minimum, and blank as the maximum.
7. If an answer is in the format of "<1", we take 1 as the maximum, and 0 as the minimum.
8. If the answer does not match the question, we treat it as a blank.
9. If an answer is "no", we assume the numeric response is 0. If an answer is "yes" and the numeric response is blank, we assume the response is blank. If an answer is "yes" and the numeric response is 0, we assume the response is 0.
10. Based on deposition testimony, some drivers' survey responses were updated. Kenneth Nevarez answered "no" for some questions and then provided numeric answers. We use those numbers since he stated in his deposition that he intended to answer "yes". John Rivero answered "no" for 11 and 12, and then provided numeric answers. We take "no" as answers and treat the numbers as zeros since according to his deposition, he intended to answer "no".

**Figure 11**  
**Statistical Tests of Survey Responses**  
**Apple Valley vs Red Bluff**

Question	Difference of Means (A)-(B)	p-Value	Percentage Difference of Means	Apple Valley		Red Bluff	
				Number of Responses	Mean (A)	Number of Responses	Mean (B)
3a Usual pretrip length	0.64	0.78	4.8%	11	13.91	15	13.27
4a Usual post-trip length	0.57	0.78	5.4%	11	11.00	15	10.43
5a # roadside/weigh inspections in usual month	-2.13 *	0.07	443.9%	8	0.48	11	2.61
5b Usual inspection length	2.24	0.74	12.4%	9	20.28	13	18.04
6a # washings in 10 usual trips	0.03	0.97	1.4%	9	2.17	11	2.14
6b Usual washing length	7.48	0.27	27.4%	11	34.82	15	27.33
7a # weighings outside Wal-Mart in 10 usual trips	1.06	0.19	85.0%	8	2.31	10	1.25
7b Usual length of weighing outside Wal-Mart	2.14	0.70	9.9%	11	23.64	15	21.50
8a # adjustments outside Wal-Mart in 10 usual trips	0.28	0.46	31.3%	9	1.17	9	0.89
8b Usual time to adjust	-11.71	0.45	43.5%	9	26.94	13	38.65
9a # refuelings at Wal-Mart in 10 usual trips	-1.10	0.43	21.9%	10	5.05	13	6.15
9b Usual time to refuel at Wal-Mart	-0.77	0.80	4.6%	11	16.73	15	17.50
10a # refuelings outside Wal-Mart in 10 usual trips	-0.15	0.88	9.4%	10	1.60	10	1.75
10b Usual time to refuel outside Wal-Mart	-5.75	0.25	34.3%	10	16.75	15	22.50
11a # meetings at start of day or trip in 10 usual trips	-0.52	0.63	6.7%	10	7.75	13	8.27
11b Usual length of meeting at start of day or trip	1.45	0.76	16.1%	10	10.45	13	9.00
12a # meetings at end of trip in 10 usual trips	-0.37	0.80	5.4%	10	6.80	12	7.17
12b Usual length of meeting at end of trip	6.05 ***	0.01	139.3%	10	10.40	13	4.35
13a # times waiting for loading in 10 usual trips	0.75	0.32	42.9%	9	2.50	10	1.75
13b Usual time waiting for loading	2.56	0.89	3.4%	8	76.88	11	74.32
14a # times waiting for unloading in 10 usual trips	1.56 *	0.06	66.7%	9	3.89	9	2.33
14b Usual time waiting for unloading	-13.79 *	0.09	35.8%	10	38.50	12	52.29
15a # rest breaks in 10 usual trips	-8.51 **	0.01	156.1%	10	5.45	12	13.96
15b Usual rest break length	0.75	0.68	6.5%	10	12.25	14	11.50

**Sources:**

1. 2016 Plaintiff Questionnaires

**Notes:**

1. The means reported are the averages of survey responses for which a mean could be computed, i.e., responses with a single number or a range of numbers.
2. \* denotes significance at the 10% level, \*\* denotes significance at the 5% level, and \*\*\* denotes significance at the 1% level.
3. The number of responses per domicile to a given question may be less than the number of survey respondents domiciled at that location to the extent that some drivers did not respond, or provided a non-numeric response, to that question.
4. If an answer is vague with a number, such as "60 depends on...", then we take the number as the answer.
5. If an answer is nonnumeric such as "varies", "do not recall", "unknown", etc., we treat it as a blank.
6. If an answer is in the format of ">#" or "#+", we take # as the minimum, and blank as the maximum.
7. If an answer is in the format of "<1", we take 1 as the maximum, and 0 as the minimum.
8. If the answer does not match the question, we treat it as a blank.
9. If an answer is "no", we assume the numeric response is 0. If an answer is "yes" and the numeric response is blank, we assume the response is blank. If an answer is "yes" and the numeric response is 0, we assume the response is 0.
10. Based on deposition testimony, some drivers' survey responses were updated. Kenneth Nevarez answered "no" for some questions and then provided numeric answers. We use those numbers since he stated in his deposition that he intended to answer "yes". John Rivero answered "no" for 11 and 12, and then provided numeric answers. We take "no" as answers and treat the numbers as zeros since according to his deposition, he intended to answer "no".

**Figure 12**  
**Statistical Tests of Survey Responses**  
**Apple Valley vs Porterville**

Question	Difference of Means (A)-(B)	p-Value	Percentage Difference of Means	Apple Valley		Porterville	
				Number of Responses	Mean (A)	Number of Responses	Mean (B)
3a Usual pretrip length	-1.09	0.61	7.8%	11	13.91	13	15.00
4a Usual post-trip length	-1.50	0.56	13.6%	11	11.00	13	12.50
5a # roadside/weight inspections in usual month	-2.72	0.30	566.7%	8	0.48	6	3.19
5b Usual inspection length	-6.31	0.35	31.1%	9	20.28	11	26.59
6a # washings in 10 usual trips	-1.83 *	0.09	84.6%	9	2.17	10	4.00
6b Usual washing length	3.91	0.65	12.6%	11	34.82	11	30.91
7a # weighings outside Wal-Mart in 10 usual trips	-0.41	0.67	17.7%	8	2.31	9	2.72
7b Usual length of weighing outside Wal-Mart	1.55	0.80	7.0%	11	23.64	12	22.08
8a # adjustments outside Wal-Mart in 10 usual trips	-1.17 *	0.07	100.0%	9	1.17	9	2.33
8b Usual time to adjust	3.99	0.73	17.4%	9	26.94	11	22.95
9a # refuelings at Wal-Mart in 10 usual trips	-0.33	0.82	6.6%	10	5.05	13	5.38
9b Usual time to refuel at Wal-Mart	1.31	0.69	8.5%	11	16.73	12	15.42
10a # refuelings outside Wal-Mart in 10 usual trips	-0.45	0.73	27.8%	10	1.60	11	2.05
10b Usual time to refuel outside Wal-Mart	-6.33	0.29	37.8%	10	16.75	13	23.08
11a # meetings at start of day or trip in 10 usual trips	-0.87	0.49	11.2%	10	7.75	13	8.62
11b Usual length of meeting at start of day or trip	-3.28	0.41	31.4%	10	10.45	13	13.73
12a # meetings at end of trip in 10 usual trips	1.34	0.40	24.5%	10	6.80	13	5.46
12b Usual length of meeting at end of trip	-1.14	0.78	10.9%	10	10.40	13	11.54
13a # times waiting for loading in 10 usual trips	-1.85 *	0.10	74.0%	9	2.50	10	4.35
13b Usual time waiting for loading	10.13	0.68	15.2%	8	76.88	10	66.75
14a # times waiting for unloading in 10 usual trips	1.39	0.15	55.6%	9	3.89	11	2.50
14b Usual time waiting for unloading	6.23	0.49	19.3%	10	38.50	11	32.27
15a # rest breaks in 10 usual trips	-3.74	0.21	68.7%	10	5.45	13	9.19
15b Usual rest break length	-1.92	0.48	15.6%	10	12.25	12	14.17

**Sources:**

1. 2016 Plaintiff Questionnaires

**Notes:**

1. The means reported are the averages of survey responses for which a mean could be computed, i.e., responses with a single number or a range of numbers.
2. \* denotes significance at the 10% level, \*\* denotes significance at the 5% level, and \*\*\* denotes significance at the 1% level.
3. The number of responses per domicile to a given question may be less than the number of survey respondents domiciled at that location to the extent that some drivers did not respond, or provided a non-numeric response, to that question.
4. If an answer is vague with a number, such as "60 depends on...", then we take the number as the answer.
5. If an answer is nonnumeric such as "varies", "do not recall", "unknown", etc., we treat it as a blank.
6. If an answer is in the format of ">#" or "#+", we take # as the minimum, and blank as the maximum.
7. If an answer is in the format of "<1", we take 1 as the maximum, and 0 as the minimum.
8. If the answer does not match the question, we treat it as a blank.
9. If an answer is "no", we assume the numeric response is 0. If an answer is "yes" and the numeric response is blank, we assume the response is blank. If an answer is "yes" and the numeric response is 0, we assume the response is 0.
10. Based on deposition testimony, some drivers' survey responses were updated. Kenneth Nevarez answered "no" for some questions and then provided numeric answers. We use those numbers since he stated in his deposition that he intended to answer "yes". John Rivero answered "no" for 11 and 12, and then provided numeric answers. We take "no" as answers and treat the numbers as zeros since according to his deposition, he intended to answer "no".

#### D. REST BREAKS

55. Plaintiffs allege that it was impossible for class members to take paid rest breaks under Wal-Mart's pay system; thus they say Wal-Mart did not pay minimum wage for those rest breaks. The randomly selected class members who gave deposition testimony seem to contradict these allegations. To my knowledge, all of the randomly selected class members who were asked about it at deposition said that they were free to take rest breaks whenever they chose, with the exception that some believed that they had to take rest breaks at least every 4 hours.<sup>14</sup> Many testified that they took rest breaks while they were being paid for some other task. Many of these other tasks such as unscheduled time are paid by the minute. Others were paid through activity pay such as live load or live unload when a trailer was being loaded or unloaded and the driver himself was doing no work. According to Wal-Mart and class member testimony, other such activities potentially include waiting and thus also afford the opportunity to take a break while being paid. Appendix 2 lists instances in which randomly selected class members testified that they were free to take rest breaks at their convenience or that they took rest breaks simultaneously as they earned activity pay. These class members represent 97.5% of the randomly selected class members who were asked whether they took paid breaks, and it represents 77.5% of the randomly selected class members who were surveyed. The only one of the randomly selected class members who testified that he was not free to take rest breaks at his convenience testified in a way that suggested that the timing and duration of his rest breaks were the result of his own confusion about trucking safety laws.<sup>15</sup> This class member testified that he took rest breaks at his discretion prior to approximately 2008 or 2009, but after that he had to take a 15 minute rest break every 4 hours to comply with a law that was somehow related to Mothers Against Drunk Driving. No other deponents have testified to there being any such law. To the extent the surveyed class members are truly representative of the class, the survey responses and deposition testimony indicate that all or virtually all class members took rest breaks at their convenience and that a large but indeterminate percentage took paid rest breaks. The indeterminacy of the percentage of class members who took paid rest breaks is due to the fact that not all survey respondents were asked whether they took paid breaks and the potential that there remains a legal dispute about whether a rest break during a hook, arrive or other fixed-rate activity can constitute a paid break.

<sup>14</sup> See, e.g., Depositions of Raymond Byer p. 134, Kevin Putnam p. 69 and John Rivero p. 39.

<sup>15</sup> See Deposition of John Rivero, pp. 39-44.

At a minimum, the survey respondents' testimony suggests that a large percentage of class members took paid rest breaks. The survey and testimony give no indication that any particular percentage of class members, let alone most, were not given the opportunity to take paid rest.

#### **E. CONCURRENT ACTIVITIES**

56. According to deposition testimony from the randomly selected sample of class members, drivers engaged in many of the activities at issue in the case concurrently with each other and concurrently with other activities for which they were paid. Class members testified about doing paperwork, taking rest breaks, cleaning their trucks and fueling their trucks during paid waiting time.<sup>16</sup> They took paid meal breaks during the time that they were waiting at vendors or Wal-Mart stores. They left their trucks and performed personal activities as their trucks were being washed.<sup>17</sup>

57. Appendix 3 lists each of the randomly selected class members and identifies deposition transcript passages where they testified that they performed activities at issue in this case simultaneously as paid activities or simultaneously with other activities at issue. Every one of the randomly selected class members who filled out a survey and gave deposition testimony is listed on Appendix 3 (except for the class member who terminated from Wal-Mart immediately after hire) because they all testified to having engaged in activities concurrently. Approximately 95% testified to having engaged in the activities at issue concurrently while being paid for something else. Approximately 87% testified to having engaged in two or more different activities at issue in this case simultaneously. These percentages may be understated because not all of the randomly selected class members were asked about both types of concurrence. The fact that a deponent listed in Appendix 3 is not identified as having concurrently engaged in an activity at issue in this case while also being paid for something else does not necessarily mean that he did not engage in such activities concurrently. Rather, it may only mean he was not asked about it at deposition. Similarly, the fact that a deponent listed in Appendix 3 is not identified as having concurrently engaged in more than one of the activities at issue in this case does not necessarily mean that he did not engage in those types of activities concurrently either.

<sup>16</sup> See, e.g., Deposition of Mark Alumbaugh pp. 29-30, Deposition of Robert Garcia pp. 127-129.

<sup>17</sup> See, e.g., Deposition of Donald Auken, 02/23/16, p. 82; Deposition of Oscar Montoya, 03/08/16, p. 86; Deposition of Daniel Steele, 02/18/16, pp. 50, 54.



58. Every survey respondent/deponent who was asked at deposition testified to spending some of his layover time off-duty, away from his truck apparently free of work responsibilities. Appendix 4 identifies transcript passages for each of the randomly selected class members, other than the one deponent who terminated virtually immediately after hire, wherein the class member testifies to having spent layover time away from his truck or in one instance to his belief that he was free to do whatever he wanted during a layover. This last deponent did not testify to having actually spent time away from the truck during layover, but he was not asked that question directly. Most respondent/deponents who were asked testified either that they merely had to notify Wal-Mart if they chose to spend their entire layover time away from their trucks or else that they did not even have to do that. Appendix 4 lists respondents/deponents who gave such testimony and the transcript citations identifying where they gave it. To the extent that the respondent/deponents are representative of the class, Appendix 4 indicates that at least 63% of class members felt that they had the discretion to spend their layovers away from their trucks without needing Wal-Mart's approval to do so. This percentage is the ratio of persons testifying either that they spent layovers away from their trucks without getting permission or that they believed that they could do so (25) to the total number of surveyed, randomly selected class members (40).<sup>18</sup> Other class members gave ambiguous testimony that is consistent with them also disbelieving that Wal-Mart restricted them from taking layovers wherever they chose and a third group was not asked.<sup>19</sup> Only a small percentage of deponents gave unambiguous testimony that they thought they needed Wal-Mart's permission to sleep away from their trucks during layovers and that seeking such permission was more than a courtesy or formality.

59. To my knowledge there are no data in the discovery record isolating the time spent on purportedly uncompensated activities separate and apart from time that was being compensated for some other reason. Similarly, to my knowledge there are no data in the discovery record that would allow calculation of purportedly uncompensated time without double counting time spent on more than one purportedly uncompensated activity simultaneously. To my knowledge there are no data in the discovery record concerning the number of days that class members worked

---

<sup>18</sup> The 25 class members are Allen, Andrews, Auker, Barker, Batham, Benavidez, Brittan, Byer, Evans, Garcia, Harris, Laurye, McCulley, McFall, Nettles, Powers, Putnam, Rand, Rivero, Robinson, Ryan, Steele, Temple, Trevino, and Vasquez.

<sup>19</sup> For example, Bruce McKee and Oscar Montoya testified that they would call Wal-Mart both for permission and to let Wal-Mart know if they were taking layovers away from their trucks.

without taking at least 10 minutes of paid rest for each four hours worked or major fraction thereof. There appear to be electronic data available concerning time that class members may have spent in their trucks during layovers, but these data exist only for a few years of the damages period and there has not been any discovery from knowledgeable Wal-Mart staff explaining these data or verifying that they are reliable for that purpose. These various data limitations regarding concurrent activities, incidence of paid rest breaks and layover time spent away from Wal-Mart trucks and, to my knowledge, undisputedly outside of Wal-Mart's control persist despite the Plaintiffs' survey.

60. Plaintiffs' survey includes directions presumably intended to account for concurrent activities, but the language is ambiguous and survey respondents did not limit their survey responses to instances in which they only engaged in purportedly uncompensated activities without doing anything else. I discuss this topic further below.

61. Given the lack of information in Wal-Mart's records or Plaintiffs' survey results about concurrent activities, I am unaware of any reasonably reliable method to ascertain how much uncompensated work-time the class suffered or what particular fraction of the class members suffered any uncompensated work-time at all. Similarly, I am unaware of any method to ascertain how many days the class worked with or without taking at least ten minutes of paid rest for every four hours worked or major fraction thereof. Finally, I am unaware of any method to ascertain how many hours the class logged in their trucks during layovers while arguably under Wal-Mart's control or the hourly wages the class actually earned for these hours.

#### **IV. PLAINTIFFS' SURVEY AND SURVEY RESULTS**

##### **A. PLAINTIFFS' SURVEY INSTRUMENT**

62. Plaintiffs sent out subpoenas to purportedly randomly selected class members to sit for deposition. When class members arrived at the deposition in response to their subpoenas, they were told to complete a seven page survey regarding their employment experiences as Private Fleet Drivers for Wal-Mart. Immediately after completing the survey, each respondent gave deposition testimony about his survey responses and about his work experiences more generally. The depositions were not uniform. Questions varied from deponent to deponent.

63. A high proportion of subpoenaed class members failed to respond to the subpoenas. Plaintiffs had to attempt to serve 104 class members in order to reach 40 survey respondents. Over 60 percent of the randomly selected sample failed to respond. As discussed below, this is significant because it means that the final group of survey respondents is not a randomly selected group of class members. Although it is possible that there was no statistical relationship between likelihood of responding to the subpoena and work experiences at Wal-Mart, this condition is unlikely and there is no logical basis to assume that it is true. As I discuss, there is some empirical evidence that the survey respondents reported higher frequencies of purportedly uncompensated activities and longer durations of such than the average class member would have reported if the entire class had been surveyed.

64. Plaintiffs' survey instrument was a seven page questionnaire that included fifteen numbered questions, most of which included additional subparts. Including subparts, the questionnaire asked 39 questions. After two questions to establish that the respondent worked as a Private Fleet Driver for Wal-Mart in California during the class period, the survey instructs the respondent about how to interpret the word "trip" in subsequent questions which ask how often they performed certain activities or experienced certain events. The survey directs the respondent to "consider a TRIP as driving from one distribution center to another (or perhaps returning to the original) distribution center."<sup>20</sup> On its face this instruction is ambiguous as it instructs the respondent to define a trip in one of two ways without giving instructions about how to choose between them. Worse still, "trip" is a term of art for Wal-Mart Private Fleet drivers. Drivers record their work activities based on "trips" and fill out "trip sheets." In their everyday work, "trips" do not necessarily start or end at distribution centers.

65. The survey proceeds to ask questions about whether respondents performed the activities listed below and designated a through m. If they performed the activities, the survey would ask about the frequency of the activity and the duration of the activity.

---

<sup>20</sup> Phillips, Fractor, & Company, LLC questionnaire, p. 2.

- a. Initial daily pre-trip inspection
- b. End of day post-trip inspection
- c. CHP/DOT roadside/weigh inspection
- d. Washing truck
- e. Weighing at non-Wal-Mart weigh stations
- f. Adjustments at non-Wal-Mart weigh stations
- g. Refueling at Wal-Mart distribution centers
- h. Refueling other than at Wal-Mart locations
- i. Meeting with driver coordinator at the beginning of workday and/or when starting new trip
- j. Meeting with driver coordinator at the end of a trip
- k. Waiting for truck to be loaded
- l. Waiting for truck to be unloaded
- m. Stopping driving specifically for breaks other than meal breaks

66. The survey instructs the respondents, “[w]ith regards to the activities covered below, please respond with the usual amount of time that you were performing ONLY that activity and not also performing another task.”<sup>21</sup> It is possible that Plaintiffs intended that class members were to limit their answers regarding duration to time during which they were not concurrently engaged in other activities, but that is not how survey respondents interpreted the instruction. Based on my review, seven survey respondents were asked whether they limited their answers in this way. Six answered that they had not, and one said he was not sure. Appendix 3 lists the deposition passages I have found where the question was asked whether survey respondents limited their survey answers to time spent exclusively on the task being asked about.

---

<sup>21</sup> Phillips, Fractor, & Company, LLC questionnaire p. 2.

67. For most activities, the survey asks whether the respondent “usually” engaged in the activity in his or her “usual work duties.” The exceptions are questions about pre-trip and post trip inspections and CHP/DOT roadside/weigh inspections. Rather than ask how frequently the pre- and post-trip inspections occurred, it asks whether the respondent usually performed the task in his or her usual workday. Regarding CHP/DOT roadside/weigh inspections, the survey asks how frequently the activity occurred in a “usual” month. The survey also asks respondents how long the various activities usually took or alternatively how long the usual activity took.

68. I discuss unit nonresponse in detail below, *i.e.*, not responding to the subpoena to take the survey at all. In addition, there was item nonresponse, failing to answer specific questions. Furthermore, although not technically non-responsive, respondents often answered duration and frequency questions with ranges (*e.g.*, 5-10 coordinator meetings per 10 usual trips), with words (*e.g.*, “it varies”) or symbols (*e.g.*, a question mark). When respondents were asked what they meant when they reported ranges, they did not always say that the midpoint of the range was “usual,” or even that the frequency or duration was always within the range.<sup>22</sup> Sometimes they testified that the upper-end of the range was more typical in some way, and sometimes they testified that the bottom was more typical.<sup>23</sup> Respondents were not always asked to clarify their non-specific survey answers, and their answers were not always definitive when they were asked, so it is not possible to aggregate the survey responses to calculate an average response without guessing about what the non-specific answers mean.

69. Figure 13 illustrates the degree to which the survey responses are subject to interpretation. The figure reports the mean (*i.e.*, arithmetic average) response to each frequency- or duration-related survey question depending on how responses given as a range are treated. The first data column is based on the assumption that if a respondent reported a range then the bottom of the range represents the most accurate response. The second data column limits the analysis to responses that report a single number for frequency or duration rather than a range. The third data column is based on the assumption that the top of each reported range represents

<sup>22</sup> See, *e.g.*, Deposition of David Lopez 03/03/16, pp. 109:21-110:16; Deposition of Eric Nettles 02/26/16, p. 34:13-24.

<sup>23</sup> See, *e.g.*, Deposition of Greg Ryan 02/22/16, pp. 79:8-11 and 99:7-12 (the top end of the range was out of the ordinary); Deposition of John Rivero 02/23/16, p. 105:7-10 (did not err on the side of putting more time when filling out the survey); Deposition of Bernard Johnston 03/02/16, pp. 155:23 – 156:3 (average wait time was not at the low end of the range he provided in his survey response).

**Figure 13**  
**Mean Survey Responses,**  
**Based on Treatment of Responses Given as Ranges**

Question	Assumption:			Percentage Difference Between Highest and Lowest Estimate
	Bottom of Range	Limited to Single Numbers	Top of Range	
3a Usual pretrip length	12.87	13.41	15.18	17.93%
4a Usual post-trip length	10.21	10.19	12.36	21.34%
5a # roadside/weigh inspections in usual month	1.98	2.24	2.11	6.35%
5b Usual inspection length	20.47	21.38	23.27	13.69%
6a # washings in 10 usual trips	2.68	2.81	2.83	5.82%
6b Usual washing length	27.05	31.67	34.19	26.37%
7a # weighings outside Wal-Mart in 10 usual trips	1.93	2.17	2.15	11.39%
7b Usual length of weighing outside Wal-Mart	21.05	23.33	23.55	11.88%
8a # adjustments outside Wal-Mart in 10 usual trips	1.36	1.50	1.56	14.62%
8b Usual time to adjust	24.55	26.20	38.38	56.37%
9a # refuelings at Wal-Mart in 10 usual trips	5.36	5.81	5.78	7.77%
9b Usual time to refuel at Wal-Mart	14.50	16.59	19.41	33.86%
10a # refuelings outside Wal-Mart in 10 usual trips	1.72	1.82	1.87	8.86%
10b Usual time to refuel outside Wal-Mart	19.34	20.77	23.03	19.05%
11a # meetings at start of day or trip in 10 usual trips	8.14	8.21	8.36	2.73%
11b Usual length of meeting at start of day or trip	9.14	11.20	12.97	42.00%
12a # meetings at end of trip in 10 usual trips	6.29	6.29	6.57	4.55%
12b Usual length of meeting at end of trip	6.95	7.69	10.25	47.57%
13a # times waiting for loading in 10 usual trips	2.67	2.64	3.03	14.94%
13b Usual time waiting for loading	51.45	50.31	92.93	84.71%
14a # times waiting for unloading in 10 usual trips	2.77	2.81	2.93	5.94%
14b Usual time waiting for unloading	30.61	30.79	52.27	70.79%
15a # rest breaks in 10 usual trips	9.25	8.87	10.57	19.17%
15b Usual rest break length	11.44	13.00	13.75	20.15%

Source: 2016 Plaintiff Questionnaires

**Notes:**

- For each question, the "Bottom of Range" column reports the mean of drivers' minimum responses of that question; the "Limited to Single Numbers" column reports the mean of responses with one single number; and the "Top of Range" column reports the mean of drivers' maximum responses of that question.
- If an answer is vague with a number, such as "60 depends on...", then we take the number as the answer.
- If an answer is nonnumeric such as "varies", "do not recall", "unknown", etc., we treat it as a blank.
- If an answer is in the format of ">#" or "#+", we take # as the minimum, and blank as the maximum.
- If an answer is in the format of "<1", we take 1 as the maximum, and 0 as the minimum.
- If the answer does not match the question, we treat it as a blank.
- If an answer is "no", we assume the numeric response is 0. If an answer is "yes" and the numeric response is blank, we assume the response is blank. If an answer is "yes" and the numeric response is 0, we assume the response is 0.
- Based on deposition testimony, some drivers' survey responses were updated. Kenneth Nevarez answered "no" for some questions and then provided numeric answers. We use those numbers since he stated in his deposition that he intended to answer "yes". John Rivero answered "no" for 11 and 12, and then provided numeric answers. We take "no" as answers and treat the numbers as zeros since according to his deposition, he intended to answer "no".

the most accurate response. The last data column shows the percentage difference between the high and low of the three estimated averages for each question. For example, Question 13 concerned usual time waiting to load. If the average is calculated based solely on persons who reported a specific time in minutes, the survey average was 50.3 minutes. If the survey average is constructed assuming that for respondents who reported a range, the top of their range was the more accurate estimate, then the mean would be 92.9 minutes or 85% more. Across all of the frequency and duration questions on Plaintiffs' survey, there was on average a 24% range in mean response depending on how one interpreted non-specific numerical answers. Misinterpreting respondents' answers about both frequency and duration can have a compound effect. If the frequency is overstated by 24% and the duration is also overstated by 24%, then the overall effect would be to overstate the sample average by 54% relative to what survey respondents intended.

70. It may seem intuitive that the appropriate treatment when the respondent reported a range would be to choose the midpoint. As I discussed above, the respondents did not uniformly support this interpretation when asked to clarify what they meant when reporting ranges. Moreover, there are also significant differences when average responses are calculated including midpoints of ranges or if they are calculated based solely on survey responses limited to single, specific numbers. There is a 34% difference in the average response concerning average wait time during unloading and a 44% difference in the average response concerning average wait time during loading depending on whether the midpoints of “numerical-range” answers are included in the average calculation or “numerical-range” answers are excluded altogether.<sup>24</sup> Interval answers introduce uncertainty about what the average respondent’s answer really was.

71. The survey confused respondents, and some respondents said so. Many had difficulty interpreting the definition of trip. Several were asked specifically about this, and different respondents testified to having interpreted “trips” in different ways. Some said that they interpreted “10 usual trips” to mean every two weeks; some interpreted it as every week; some interpreted it as every 10 trips as “trips” are defined in the ordinary course of business as a Private Fleet Driver and others tried to answer based on trips that end at a distribution center.<sup>25</sup>

72. The modifiers “usual” and “usually” also caused confusion. As discussed above, drivers did not have uniform experiences over time, so it is unclear how they were interpreting “usual” and “usually.”<sup>26</sup> Some tried to interpret it as some sort of “ballpark” or “guesstimated” amount of time.<sup>27</sup>

---

<sup>24</sup> Excluding “numerical-range” answers, the average responses concerning average wait times during unloading and loading are 30.8 minutes and 50.3 minutes, respectively. Including “numerical-range” answers, these averages are 41.4 minutes for unloading and 72.4 minutes for loading.

<sup>25</sup> See, e.g., Deposition of Dana Brittan, 02/22/16, p. 82:3 (two weeks); Depositions of Robert Laurye, 03/10/16, 63:22-24 and Stanley McCulley, 04/28/16, 86:15-18 (one week); Deposition of Michael Andrews, 04/21/16, 52:16-53:2 (it varies but the current norm is probably 3-4 days); Deposition of John Goodman, 02/26/16, 97:25-98:8 (did not limit “trips” to those that were from one distribution center to another); Deposition of Angel Vasquez, 05/04/16, 65:12-22 (trips end at distribution centers). Mark Alumbaugh testified that his responses to the survey reflected trips that ended at locations other than distribution centers unless the survey question specifically mentioned trips that were from a distribution center to a distribution center (Deposition of Mark Alumbaugh, 03/18/16, 58:13-19 and 67:23-68:8).

<sup>26</sup> See, e.g., Deposition of Kenneth Nevarez 03/01/16 p. 82:15-17: (“Nothing is usual when you’re driving for Wal-Mart.”) and Deposition of Mark Alumbaugh p. 50:12-22 (“there was no such thing as usual for most of us”).

<sup>27</sup> Depositions of Bobby Carr 04/12/16, pp. 69:15-70:4 and 83:12-22 (ballpark); Michael Barker 04/11/16, pp. 54:10-24 and 108:23-109:12 (guesstimate); Kevin Putnam 03/23/16, p. 61:10-24 (guesstimate); Michael Rand 02/26/16, pp. 76:4-12 and 86:18-20 (guesstimate or average).



73. “Typical” is vague and ambiguous too. Common measures of “typicality” that are used in math and statistics are the arithmetic mean (which is what people often mean by “average”), the median and the mode. The median is the number such that half of the set of numbers being evaluated are higher and half are lower. The mode is that number that is most often represented among the set of numbers being evaluated. These measures may be markedly different from each other, and they do not all have the same arithmetic qualities. For example, if a driver washed his truck 100 times in his tenure as a Private Fleet Driver, and the average washing time was 10 minutes, that would mean that he or she spent 1,000 minutes washing his or her truck in the aggregate. However, if a driver washed his truck 100 times and the median washing time was 10 minutes, that would not mean that he or she necessarily spent 1,000 minutes washing his or her truck in total. Mathematically, with a 10 minute median time and 100 washings, aggregate washing time could range from approximately 500 minutes to any arbitrarily large amount of time. Similarly, a mode of 10 minutes would not imply 1,000 minutes of aggregate washing time either. As a mathematical matter, aggregate washing time could be much more or much less than 1,000 minutes.

74. As drivers’ experiences varied over time, typical at one time in the class period, say 2005, may not be typical for other times in the class period say 2014. Consider survey respondent Eric Nettles. Mr. Nettles started driving for Wal-Mart in 2004. Mr. Nettles testified that for his first six months at Wal-Mart, he was “running wild.” This refers to a driver who has no fixed route. Mr. Nettles would drive anywhere he was needed throughout “California, Las Vegas, the local West Coast area.” After this six month period, Mr. Nettles testified that he drove local routes out of Fontana for four years, delivering general merchandise to local areas and going home every night. According to Mr. Nettles, in approximately 2010 he reverted to running wild again.<sup>28</sup> Mr. Nettles testified that the frequency and duration of various activities at issue in this case varied depending on whether he was running wild, what distribution center he was driving from and the identity of the driver coordinator.<sup>29</sup>

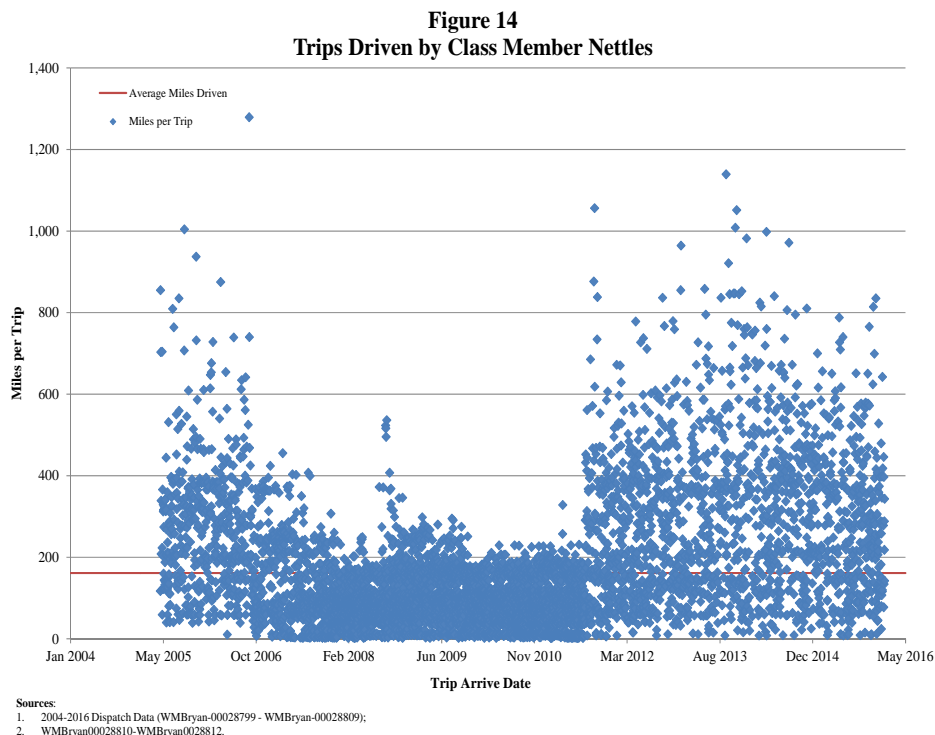
75. Figure 14 is a scatter plot of Mr. Nettles’ trips over his tenure with Wal-Mart. The figure is consistent with there being a period of time at the commencement of his employment with

<sup>28</sup> Deposition of Eric Gean Alfred Nettles, 02/26/16, pp. 18-20.

<sup>29</sup> Deposition of Eric Gean Alfred Nettles, 02/26/16, pp. 44, 50-52.



Wal-Mart when there was especially much variety in terms of trip length, followed by a multi-year period during which there was less variety, and trips most often fell within the 0 to 180 mile range. Then, a third period began during which Mr. Nettles was apparently running wild again. Figure 14 corroborates Mr. Nettles' testimony that his tenure can be divided into three separate time periods. That "the usual" apparently varied significantly among the first, second and third eras calls into question which time period Mr. Nettles speaks of when he responds to survey questions about "usual trips" or "usual workdays" relating to "any time from October 10, 2004 through October 16, 2015."<sup>30</sup>



76. Figure 14 contradicts Mr. Nettle's recollections of time. He apparently ran wild longer than six months when he started at Wal-Mart, and he ended his Fontana local runs later than 2010. Mr. Nettles' failure to accurately recollect the dates at which his employment circumstances changed radically calls into question whether Mr. Nettles' estimates of usual times and frequencies would be accurate even if he were trying to provide a figure that was "usual" for his whole tenure rather than "usual" for one of the three distinct eras. As it stands, Mr. Nettles does not always appear to have been trying to represent what was "usual" during his whole

<sup>30</sup> Plaintiffs' questionnaire p. 1.

tenure, but rather he seemed sometimes to be limiting his answer to the time when he was driving local routes in Fontana even though his time driving mostly local routes represents less than half of his tenure as a Private Fleet Driver.<sup>31</sup>

77. In addition to the confusing and ambiguous instructions regarding concurrent activities and the terms “trip,” and “usual”, Plaintiffs’ survey was unclear in other ways, for example by asking about separate and distinct events or activities as though they were the same things. The survey asks about the frequency and duration of CHP/DOT roadside/weigh inspections. Survey respondents testified that CHP/DOT roadside/weigh inspections are three different things. CHP roadside inspections were apparently quite rare, and the duration was relatively brief. DOT inspections at weigh scales were also rare, but the duration could be lengthy. Weigh inspections could be frequent, but Wal-Mart trucks were routinely waived through. This meant that they did not take much time, and drivers accrued mileage pay as they occurred.<sup>32</sup>

78. Appendix 5 illustrates some of the confusion experienced by survey respondents. For each respondent listed in Appendix 5, I have identified deposition testimony in which the respondent states some misunderstanding, ambiguity or confusion related to Plaintiffs’ survey instrument.

79. Appendix 6 further illustrates respondents’ confusion and the general unreliability of respondents’ survey answers. Each respondent listed in Appendix 6 either testified explicitly that his or her responses were unreliable, changed at least one of his or her answers at the follow-on deposition to the survey, gave inconsistent deposition testimony to his survey answers or gave conceptually inconsistent survey responses, e.g., testified that he did not usually engage in an

---

<sup>31</sup> Deposition of Eric Gean Alfred Nettles, 02/26/16, p. 53:

“When you were running a local route, how many days a week were you driving?

A. Five days a week.

Q. Got it.

And is that where this five number comes from --

A. Yes.

Q. -- on the survey?

A. Correct.”

<sup>32</sup> See Deposition of Doug Allen 03/28/16, pp. 37:17-38:10 (driving through the weigh station is a frequent occurrence but DOT inspections are rare); Deposition of Dana Brittan 02/22/16 p. 78:2-18 (Wal-Mart trucks usually get a bypass at weigh inspections, and they can keep driving through); Deposition of Bobby Carr 04/12/16, p. 63:1-16 (inspections take longer than driving through the scale system); Deposition of James Batham 05/10/16, pp. 66:22-68:9 (there are quarterly inspections and there are also random CHP inspections at the side of the road).

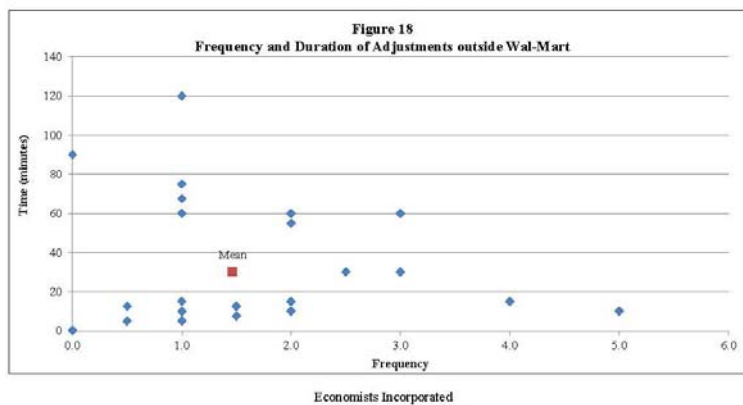
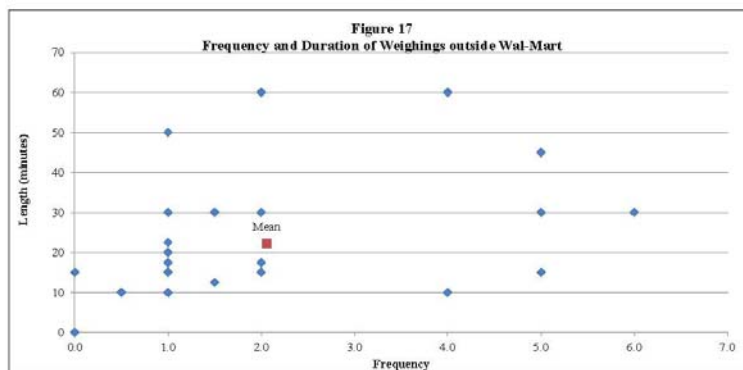
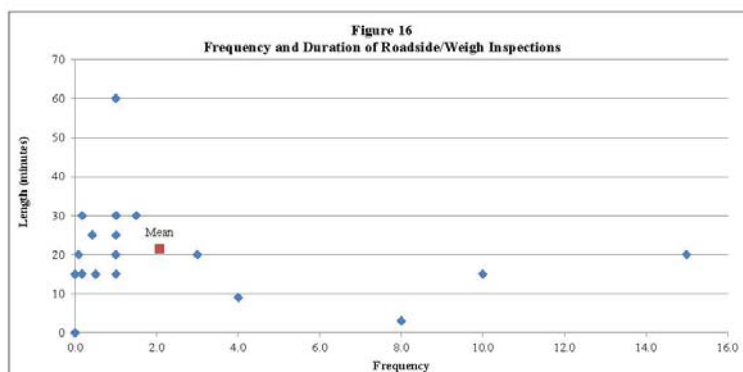
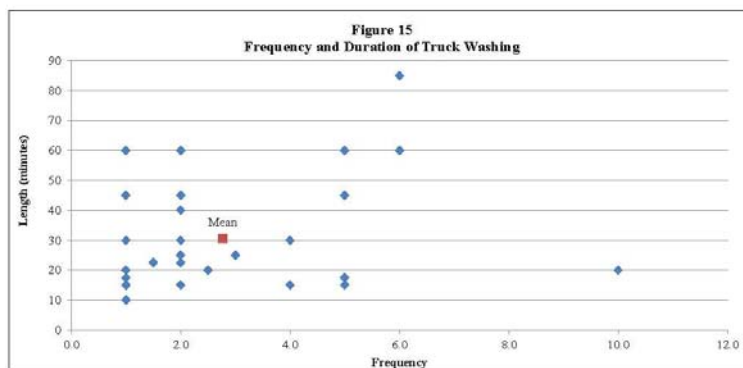
activity but also provided a non-zero number of times per 10 usual trips that he or she would engage in the activity in ten usual trips. Most of the respondents exhibited confusion in one or more objective ways.

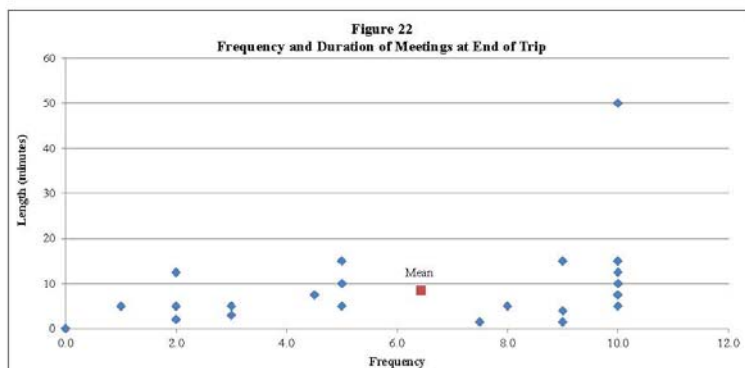
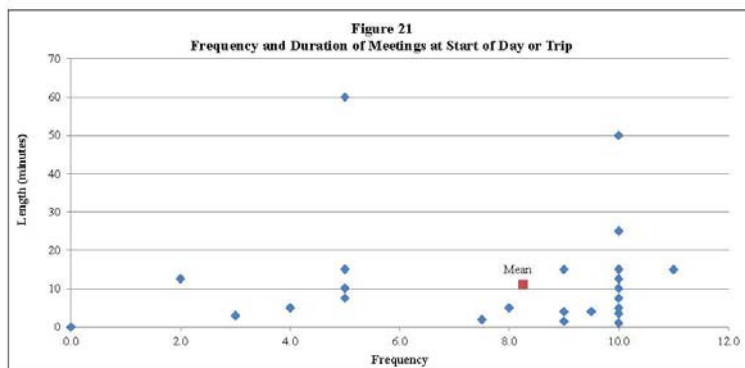
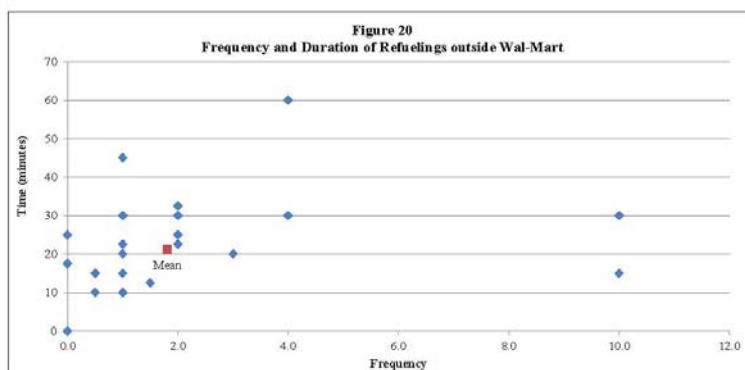
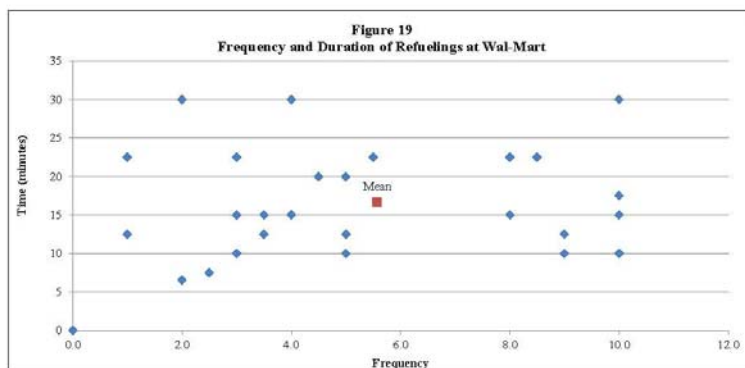
## **B. HETEROGENEITY OF REPORTED EXPERIENCES**

80. Plaintiffs' survey indicates that there is tremendous variation in the rate at which class members recall engaging in purportedly uncompensated tasks and the amount of time they recall spending on them. To the extent that the Plaintiffs' survey is valid, class members' experiences are dissimilar to each other.

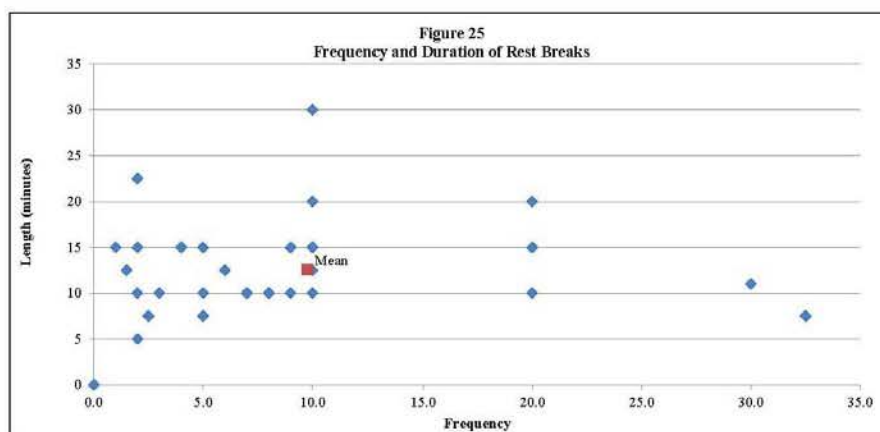
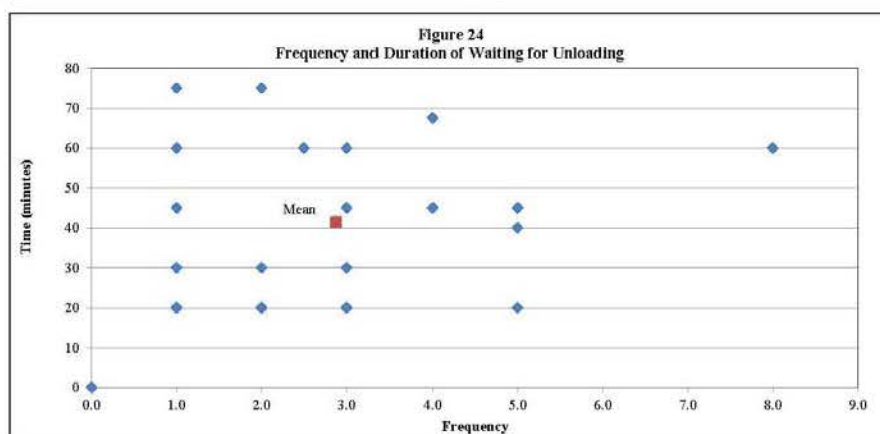
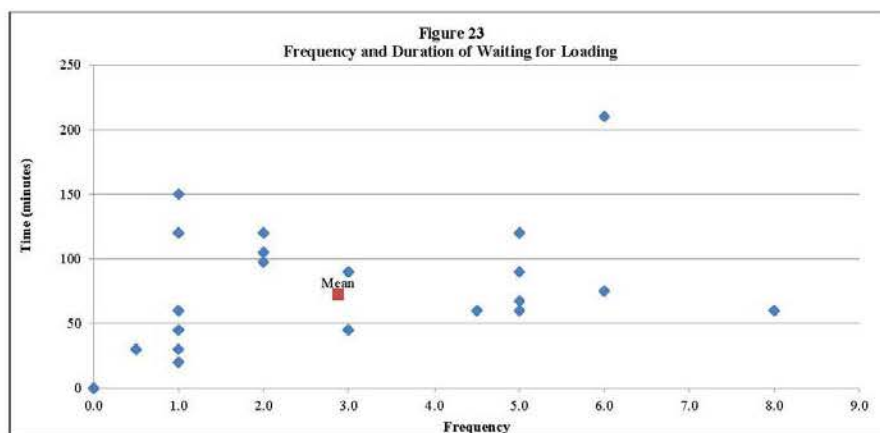
81. Figures 15 through 25 illustrate the dissimilarities among class members. Plaintiffs' survey asks respondents about the frequency and duration of 11 activities—washing trucks, roadside/weight inspections, weighing outside of Wal-Mart, adjusting weights, refueling at Wal-Mart locations, refueling outside of Wal-Mart locations, meeting at the start of certain trips, meeting at the end of certain trips, waiting for trucks to be loaded, waiting for trucks to be unloaded, and stopping driving for unpaid rest breaks. Figures 15 through 25 are scatter plots which graphically represent class members' responses to these frequency and duration questions. Each one of Figures 15 through 25 is based on respondents' answers about one of the 11 activities that Plaintiffs' survey asked about.

82. Figure 15 is based upon answers to questions about the frequency and duration of truck washing. The frequency at which respondents said they washed their trucks is measured on the x-axis and the time they say it tended to take them is measured on the y-axis. Each respondent who answered the questions regarding frequency and duration of truck washing with a numerical answer is represented on Figure 15 by a diamond corresponding at the frequency and duration that corresponds to that class member's answers. The red square on Figure 15 corresponds to the average frequency and average time across all respondents. If class members were similar in terms of the frequency at which they say truck washing occurred and the length of time that a washing typically took, then all of the diamonds on Figure 15 should be clustered in close proximity to the red square. If frequencies were the same or quite similar, but durations varied, the blue diamonds would be clustered on a vertical line that runs through the red square. If the durations were the same or quite similar, but frequencies varied, the blue diamonds would be on or close to a horizontal line that runs through the red square. However, rather than being





Economists Incorporated



Sources for Figure 15 - 25:  
1. 2016 Plaintiff Questionnaires

Notes for Figure 15 - 25:

1. If an answer is vague with a number, such as "60 depends on...", then we take the number as the answer. If an answer is nonnumeric, such as "varies", we treat it as a blank. If an answer is in the format of ">#\" or "#+\"", we take # as the minimum, and blank as the maximum. If an answer is in the format of "<#\", we take 1 as the maximum, and 0 as the minimum. If the answer does not match the question, we treat it as a blank. If an answer is "no", we assume the numeric response is 0. If an answer is "yes" and the numeric response is blank, we assume the response is blank.
2. Based on deposition testimony, some drivers' survey responses were updated. Kenneth Nevarez answered "no" for some questions and then provided numeric answers. We use those numbers since he stated in his deposition that he intended to answer "yes". John Rivero answered "no" for 11 and 12, and then provided numeric answers. We take "no" as answers and treat the numbers as zeros since according to his deposition, he intended to answer "no".

Economists Incorporated

clustered close to the average, the diamonds on Figure 15 are scattered about somewhat haphazardly. Observations (datapoints) are not all on or close to any vertical line because respondents reported dissimilar frequencies at which they washed their trucks. Observations are not on or near a horizontal line because respondents reported dissimilar times for a typical truck washing. Figure 15 shows that the frequency and duration of truck washing was quite different from respondent to respondent. At one extreme, several respondents reported washing their trucks only once per ten usual trips. At the other extreme, one respondent reported washing his truck ten times as often as that. One respondent reported that he “usually” spent 10 minutes washing his truck. Another said he “usually” spent close to an hour and a half. Figures 16 through 25 show similar heterogeneity for the other activities that Plaintiffs’ survey asked about.

83. Appendix 7 is a collection of similar scatter plots based on the deposition excerpts that Dr. Phillips relied upon instead of the survey responses. The deposition testimony illustrated similar heterogeneity of experiences across class members as the survey responses did.

84. Figure 26 further illustrates the dissimilarities among class members. Plaintiffs’ survey asks respondents whether they usually or sometimes engaged in various activities. If so, it asks how frequently they engaged in those activities and how long each activity would usually take. For respondents who reported engaging in an activity at all, Figure 26 shows the range of reported frequencies and the range of reported durations across respondents. For example, among respondents who answered yes when asked “did you usually perform a first, daily, pre-trip inspection of your tractor and/or trailer at the beginning of your usual workday,” the minimum amount of time that a respondent reported was 5 minutes while the maximum was 30 minutes, a 500% difference. Measured in percentage terms, the smallest range of difference in answers was 400%. This 400% range related to the frequency of adjustments at non-Wal-Mart weigh stations. The differences across respondents in reported frequencies and durations were remarkable when viewed either in absolute or percentage terms. For example, the range of answers regarding frequency of stops specifically for rest breaks per 10 usual trips ranged from 1 to 50. The reported usual waiting time for loading in 10 usual trips ranged from 10 minutes to 4 hours. As extreme as the variations reported on Figure 26 are, they *understate* variation in class members’ responses. Figure 26 is limited to people who reported engaging in the relevant activities at all. For each activity except for pre-trip inspections and washing the truck, at least one respondent said that the activity did not occur at all in 10 usual trips. For example, although one respondent

claimed that he usually had to wait for his truck to be unloaded in 8 out of 10 usual trips, roughly 10% of survey respondents said they did not usually have to wait for their trucks to be unloaded in any of 10 usual trips.

**Figure 26**  
**Ranges of Reported Frequencies and Durations Among Survey Responses**

Question	Minimum	Maximum	Range	Range as % of Minimum
3a Usual pretrip length	5	30	25	500%
4a Usual post-trip length	5	30	25	500%
5a # roadside/weigh inspections in usual month	0.1	15	14.9	17900%
5b Usual inspection length	3	60	57	1900%
6a # washings in 10 usual trips	1	10	9	900%
6b Usual washing length	6	150	144	2400%
7a # weighings outside Wal-Mart in 10 usual trips	1	6	5	500%
7b Usual length of weighing outside Wal-Mart	5	60	55	1100%
8a # adjustments outside Wal-Mart in 10 usual trips	1	5	4	400%
8b Usual time to adjust	5	180	175	3500%
9a # refuelings at Wal-Mart in 10 usual trips	1	10	9	900%
9b Usual time to refuel at Wal-Mart	5	45	40	800%
10a # refuelings outside Wal-Mart in 10 usual trips	0.5	10	9.5	1900%
10b Usual time to refuel outside Wal-Mart	10	60	50	500%
11a # meetings at start of day or trip in 10 usual trips	2	11	9	450%
11b Usual length of meeting at start of day or trip	1	90	89	8900%
12a # meetings at end of trip in 10 usual trips	1	10	9	900%
12b Usual length of meeting at end of trip	1	90	89	8900%
13a # times waiting for loading in 10 usual trips	1	10	9	900%
13b Usual time waiting for loading	10	240	230	2300%
14a # times waiting for unloading in 10 usual trips	1	8	7	700%
14b Usual time waiting for unloading	15	120	105	700%
15a # rest breaks in 10 usual trips	1	50	49	4900%
15b Usual rest break length	5	30	25	500%

**Source:** 2016 Plaintiff Questionnaires

**Notes:**

1. The table presents the non-zero minimum, maximum, and range for each numeric response across all drivers.
2. If an answer is vague with a number, such as "60 depends on...", then we take the number as the answer.
3. If an answer is nonnumeric such as "varies", "do not recall", "unknown", etc., we treat it as a blank.
4. If an answer is in the format of ">#" or "#+", we take # as the minimum, and blank as the maximum.
5. If an answer is in the format of "<1", we take 1 as the maximum, and 0 as the minimum.
6. If the answer does not match the question, we treat it as a blank.
7. If an answer is "no", we assume the numeric response is 0. If an answer is "yes" and the numeric response is blank, we assume the response is blank. If an answer is "yes" and the numeric response is 0, we assume the response is 0.
8. Based on deposition testimony, some drivers' survey responses were updated. Kenneth Nevarez answered "no" for some questions and then provided numeric answers. We use those numbers since he stated in his deposition that he intended to answer "yes". John Rivero answered "no" for 11 and 12, and then provided numeric answers. We take "no" as answers and treat the numbers as zeros since according to his deposition, he intended to answer "no".

85. I am unaware of a reliable method to decompose the extreme variation across class members in the reported frequency and duration of various activities into variation that is attributable to recall error, variation that is attributable to differences in how respondents interpreted the questions and actual variation in frequencies and durations of frequencies and durations across class members. However, taken at face value, Plaintiffs' survey indicates that there are major variations in the time respondents spent on the tasks that Plaintiffs say were not



compensated. To the extent that the randomly selected class members are representative of the class and that Plaintiffs' survey is valid, Plaintiffs' survey indicates significant heterogeneity across the class in terms of the frequency and duration of the tasks at issue in this case.

### C. SAMPLING ERROR

86. Plaintiffs' survey suffers from very high sampling error. Sampling error is divergence between the average response for sample respondents and the average response that would be observed if the entire population were sampled, in this case the entire class. Sampling error is often discussed in terms of "margin of error." Margin of error describes the range of values that the researcher is 95% confident includes the true population average. It is often expressed as the sample average plus or minus some percentage. For example, if a sample of 1000 randomly selected voters were asked who they intend to vote for in an upcoming election, and 55% of sampled voters say they will vote for the Republican candidate with a margin of error of 5%, then the researcher can be 95% confident that 50% to 60% of voters would say they plan to vote Republican if all voters were surveyed rather than only 1,000. The best estimate is 55% exactly, but the likelihood of that estimate being exactly right is low. Rather, the researcher can be confident at the 95% level that the true answer lies within the margin of error, 55% plus or minus 5%.

87. The margin of error is a function of a) sample size and b) variation among the population in the dimension of interest. Generally speaking, increasing the sample size will reduce the margin of error. At the same time, the less variation there is among the underlying population in the relevant dimension, the smaller the sample size is needed to reduce the margin of error to an acceptable percentage. In the case at hand, the high variation across respondents (*e.g.*, responses of anywhere from 1 to 50 unpaid rest breaks per 10 usual trips) yields very high margins of error. Leaving aside the other flaws, biases and errors to be discussed below, random chance undermines confidence in Plaintiffs' survey responses as accurately reflecting the experiences of class members on the whole.

88. Figure 27 below illustrates how large the margins of error are related to Plaintiffs' survey results. Figure 27 is calculated based on survey respondents' specific numerical answers to questions about duration and frequency. In other words, if a respondent answered a question with a range or words or not at all, his or her answer to that question is omitted from Figure 27.

89. Figure 27 shows that the margin of error for Plaintiffs' survey ranged from 11.8% for the question about usual, first, daily pre-trip inspection length to 72.0% regarding the number of CHP/DOT roadside/weight inspections in a usual month. The average margin of error across all frequency and duration questions was 29.1%. Leaving aside the other irreparable flaws in Plaintiffs' survey and assuming *arguendo* that the survey were methodologically sound, minimum-wage-related damages estimates based on Plaintiffs' survey results would be subject to a large amount of random error. For example, if the average frequency of a task were overstated by 29.1% due to random error, and the average duration of the task were also overstated by 29.1% due to random error, then total time spent on the task would be overstated by approximately two-thirds due to random error alone ( $129.1\% \times 129.1\% = 166.7\%$ ). Figure 28 illustrates the potential cumulative effect of overstating both the frequency and duration of the tasks at issue due solely to random error. For each of the purportedly uncompensated tasks about which Plaintiffs' survey asks both frequency and duration, I report the compound error if both the true class-average frequency and true class-average duration were at the bottom of the 95% confidence intervals derived from the sample. As shown in Figure 28, random error may result in aggregate damages being overstated by 46% to 112% for each activity due solely to random error. Individualized damages estimates would be off by even more due to the lack of uniformity across the class. This random error is additive to the interpretation error discussed above. That is, Figure 28 calculates the sample average by ignoring respondents who answered in averages. As discussed, this assumption introduces additional potential error besides the random, sampling error illustrated by Figure 28. This random error illustrated in Figure 28 also excludes additional error related to systematic bias that I discuss in the sections that follow.

**Figure 27**  
**Survey Response Margins of Error**  
**Based on Specific Numerical Responses to Survey Questions**

Question	Mean	# Numeric Responses	Lower 95% Confidence Interval	Upper 95% Confidence Interval	Margin of Error (as % of Mean)
3a Usual pretrip length	13.41	27	11.82	14.99	11.8%
4a Usual post-trip length	10.19	27	7.83	12.54	23.1%
5a # roadside/weigh inspections in usual month	2.24	22	0.63	3.85	72.0%
5b Usual inspection length	21.38	24	16.37	26.38	23.4%
6a # washings in 10 usual trips	2.81	27	1.96	3.66	30.2%
6b Usual washing length	31.67	27	25.32	38.01	20.0%
7a # weighings outside Wal-Mart in 10 usual trips	2.17	23	1.46	2.89	32.9%
7b Usual length of weighing outside Wal-Mart	23.33	30	17.95	28.72	23.1%
8a # adjustments outside Wal-Mart in 10 usual trips	1.50	22	0.96	2.04	36.2%
8b Usual time to adjust	26.20	25	14.48	37.92	44.7%
9a # refuelings at Wal-Mart in 10 usual trips	5.81	27	4.52	7.11	22.3%
9b Usual time to refuel at Wal-Mart	16.59	22	13.10	20.09	21.1%
10a # refuelings outside Wal-Mart in 10 usual trips	1.82	28	0.89	2.76	51.3%
10b Usual time to refuel outside Wal-Mart	20.77	26	15.37	26.17	26.0%
11a # meetings at start of day or trip in 10 usual trips	8.21	33	7.21	9.21	12.2%
11b Usual length of meeting at start of day or trip	11.20	25	6.54	15.86	41.6%
12a # meetings at end of trip in 10 usual trips	6.29	31	4.89	7.69	22.3%
12b Usual length of meeting at end of trip	7.69	26	5.52	9.87	28.3%
13a # times waiting for loading in 10 usual trips	2.64	25	1.79	3.49	32.0%
13b Usual time waiting for loading	50.31	16	36.88	63.75	26.7%
14a # times waiting for unloading in 10 usual trips	2.81	27	2.05	3.58	27.1%
14b Usual time waiting for unloading	30.79	19	22.50	39.08	26.9%
15a # rest breaks in 10 usual trips	8.87	31	6.63	11.11	25.2%
15b Usual rest break length	13.00	25	10.81	15.19	16.9%

Source: 2016 Plaintiff Questionnaires

**Notes:**

1. The values reported are the average of responses that were given as a specific number and not as a range.
2. If an answer is vague with a number, such as "60 depends on...", then we take the number as the answer.
3. If an answer is nonnumeric such as "varies", "do not recall", "unknown", etc., we treat it as a blank.
4. If an answer is in the format of ">#" or "#+", we take # as the minimum, and blank as the maximum.
5. If an answer is in the format of "<#", we take 1 as the maximum, and 0 as the minimum.
6. If the answer does not match the question, we treat it as a blank.
7. If an answer is "no", we assume the numeric response is 0. If an answer is "yes" and the numeric response is blank, we assume the response is blank. If an answer is "yes" and the numeric response is 0, we assume the response is 0.
8. Based on deposition testimony, some drivers' survey responses were updated. Kenneth Nevarez answered "no" for some questions and then provided numeric answers. We use those numbers since he stated in his deposition that he intended to answer "yes". John Rivero answered "no" for 11 and 12, and then provided numeric answers. We take "no" as answers and treat the numbers as zeros since according to his deposition, he intended to answer "no".

**Figure 28**  
**Potential Overstatement Due to Random Error**

Activity	Margin of Error - Frequency	Margin of Error - Duration	Compound Error
Roadside/weigh inspections	72.0%	23.4%	112.2%
Truck Washings	30.2%	20.0%	56.3%
Weighings outside Wal-Mart	32.9%	23.1%	63.5%
Adjustments outside Wal-Mart	36.2%	44.7%	97.2%
Refuelings at Wal-Mart	22.3%	21.1%	48.1%
Refuelings outside Wal-Mart	51.3%	26.0%	90.6%
Meetings at start of day or trip	12.2%	41.6%	58.9%
Meetings at end of trip	22.3%	28.3%	56.8%
Waiting for loading	32.0%	26.7%	67.3%
Waiting for unloading	27.1%	26.9%	61.3%
Rest breaks	25.2%	16.9%	46.3%

Source: 2016 Plaintiff Questionnaires

**Note:** The margin or error values are calculated using the average of responses that were given as a specific number and not as a range.

**V. THE UNRELIABILITY OF RECALL AS A MEASURE OF THE FREQUENCY AND DURATION OF REPEATED ACTIVITIES OR BEHAVIORS**

90. A core premise of Plaintiffs' survey and Dr. Phillips' reliance on deposition passages is that class members' recollections in 2016 about the frequency and duration of their activities dating back as far as 2004 are statistically unbiased estimates of the frequency and duration of the activities that they actually engaged in over that time. This premise is invalid for at least three reasons.

91. First, it is well known in fields that utilize self-reported survey data concerning the frequency of activities or behaviors that such data are subject to statistical bias, i.e., error that does not converge to zero as the sample size increases.<sup>33</sup> The frequency at which respondents report behaviors may be a multiple of the frequency at which the behaviors actually occurred. For any of the activities at issue in this case, there is no empirical or theoretical basis to assume that the average reported frequency from the sample (or the average or median for some selected subset of respondents) will accurately reflect the average frequency that the respondents or the class as a whole experienced. To the contrary, the survey literature indicates that respondents typically overstate the frequency at which incidents occur in circumstances such as we have here where class members did not have uniform daily experiences throughout their tenures as drivers.<sup>34</sup>

92. Second, recall error increases as the recall window expands. The recall window is the length of time over which respondents are asked to remember. Some research in the health economics field suggests that significant recall error manifests when respondents are asked to recall events more than three days past. In the case at hand, respondents were asked to recall back to their dates of hire or October 2004, whichever was later. Therefore, the recall window was approximately 11 1/2 years for many respondents. The shortest recall window for any

---

<sup>33</sup> For example, see Phillip M. Clarke, Cenil G. Fiebig and Ulf-G. Gerdtham, "Optimal Recall Length in Survey Design," *Journal of Health Economics* 27 (2008), p. 1275, "Self-reported data collected via surveys are a key input into a wide range of research conducted by economists. It is well known that such data are subject to measurement error that arises when respondents are asked to recall past utilization."

<sup>34</sup> Robert Groves, Floyd Fowler, Jr., Mick P. Couper, James Lepkowski, Eleanor Singer and Roger Tourangeau, *Survey Methodology*, Second Edition, 2009, p. 235, "[R]espondents often turn to rate based estimation when there are more than seven events or so to report. The literature suggests that rate-based estimation often leads to overestimates of behavioral frequencies. Apparently, people overestimate rates when the rate fluctuates or when there are exceptions to what usually happens."

respondent to Plaintiffs' survey was almost three years. The recall window exceeded five years for 90% of respondents who actually drove trips for Wal-Mart.

93. Third, people's perceptions of the duration of events or activities, particularly waiting time, is not an accurate measure of the actual duration of those events or activities. This error is not random either. People systematically overstate durations of passive activities, especially wait time, so that the average reported duration from a sample of survey respondents may tend to overstate the actual duration of the activity that they are reporting on. People's perceptions of the duration of an event may be a multiple of the duration that the respondents actually experienced. Taken together, these well-known facts about surveys imply that any damages analysis based on Plaintiffs' survey, or selected class members' recollections at deposition, will be subject to incalculable, potentially enormous error. These cognitive errors are in addition to sampling error discussed above and the other biases discussed below. In the sections that follow, I summarize some of the academic research demonstrating that a) self-reported surveys do not accurately measure the frequency at which respondents historically engaged in behaviors or activities, b) recall error increases over the recall window, so Plaintiffs' survey results are especially suspect and c) people overstate the duration of passive events they have experienced such as waiting.

#### **A. RECALL BIAS**

94. Recall bias in the context of surveys of self-reported behavior is inaccuracy due to memory failure. The literature in this field identifies two sources of error—omissions due to forgetting which leads to underreporting of frequencies of behaviors or activities and telescoping which leads to over-reporting of behaviors or frequencies.<sup>35</sup> Telescoping, also called boundary failure, occurs when the respondent misremembers when an activity occurred. For example, if a survey respondent were basing his recollection of how often activities occurred in "ten usual trips" by counting the number of times he remembered it occurring in the most recent "ten usual trips," telescoping would occur if the respondent included in his count instances that actually occurred in older trips. Effectively, the respondent would be giving a count based on more than

---

<sup>35</sup> See for example, Seymour Sudman and Norman Bradburn, "Effects of Time Memory Factors on Response in Surveys," *Journal of the American Statistical Association*, 68, No. 344 (December 1973), p. 805, "There are two kinds of memory error that sometimes operate in different directions. The first is forgetting an episode entirely, whether it is a purchase of a product, a trip to the doctor, a law violation, or any other act. The second kind of error is compression of time (telescoping) where the event is remembered as occurring more recently than it did."

10 trips. Although this example pertains to a specific cognitive framework that a respondent may theoretically use to answer the “ten usual trips” questions, as a theoretical matter telescoping can occur if the respondent uses other methods too. For example, if he answers based on some specific set of trips or he answers based on some time frame or he actually tries to remember every trip over the class period.

95. To my knowledge, there is no conceptual reason to expect forgetting and telescoping to balance each other out, and they do not uniformly balance each other out empirically. Many researchers have documented that surveys of self-reported behaviors are biased and that the bias can lead to significant and systematic over-reporting. This literature demonstrating that surveys based on respondents’ recall are inaccurate dates back at least to the 1950s.

96. The first published article I am aware of demonstrating that recollections reported on surveys are inaccurate is “The Memory Factor in Social Surveys,” published in 1955.<sup>36</sup> This paper noted implausible differences in responses to the British Survey of Sickness which the British government administered from 1944 through 1952. Each month a different sample of respondents were surveyed about their incidence of illness over the three preceding months. For example, in January, the government surveyed a group of respondents about their rates of illness in the preceding October, November and December. In February, the government surveyed a different sample about their rates of illness in the preceding November, December and January. In March, the government would survey a third sample about their rates of illness in the preceding December, January and February. Thus for any given month, there would be three different samples from which to measure the rate of illness in that month. For any given December, there would be an estimate based on surveys taken the following January, the following February and the following March. For large samples, if recall is accurate, there should not be any statistically significant differences between the three samples. This was not the case. Different samples varied by over 70% in terms of the estimated rate at which people suffered from colds or flu in a month. The only plausible explanation for different samples to differ by so much is that one or more of the sets of responses were inaccurate.

97. Clark, Fiebig and Gerdtham reported in the *Journal of Health Economics* on a

---

<sup>36</sup> Percy Gray, “The Memory Factor in Social Surveys,” *Journal of the American Statistical Association*, 50, No. 270 (June 1955), pp. 344-363.

comparison of survey responses about number of overnight hospital stays to actual hospital records.<sup>37</sup> Every year Statistics Sweden surveys a random sample of Swedish adults concerning a variety of topics related to living conditions in Sweden including whether the respondents had been hospitalized in the three months preceding the survey. Separately, the National Board of Health and Welfare maintains a national hospital Patient Register. Researchers linked the two and compared survey respondents' reported hospital utilization to their actual hospital utilization based on national records. Three-hundred-eighty-six respondents accurately reported that they had had an overnight hospital stay in the three months preceding the survey. There were 59 false negatives, respondents who incorrectly reported that they had not been hospitalized during the relevant time period although they actually been. There were 135 false positives, respondents who incorrectly reported that they had been hospitalized during the relevant period although they actually had not been. Twenty-six percent of reported hospital stays according to survey records were false.<sup>38</sup>

98. John Robinson and Ann Bostrom published, "The Overestimated Workweek? What Time Diary Measures Suggest," in the August 1994 issue of *Monthly Labor Review*.<sup>39</sup> Their research was based on daily diary data collected in 1965, 1975 and 1985. By comparing survey responses regarding weekly hours worked to daily activity diaries which were presumably more valid, they found that people systematically overestimated their hours worked in response to surveys such as the Current Population Survey ("CPS"). The CPS is conducted by the U.S. Bureau of the Census for the Bureau of Labor Statistics. It is the primary source for labor force statistics for the U.S. Robinson and Bostrom found that the recall bias was higher the longer the actual workweek. People who worked longer workweeks also overestimated their workweeks by more. According to Table 2 of the Robinson and Bostrom article, people who reported working 40-44 hours or more per week in responding to surveys actually worked 38.6 hours per week on average according to their daily diaries. People who reported working 75 hours or more per week in responding to surveys actually worked 54.9 hours per week on average according to their daily diaries.<sup>40</sup>

99. Robinson, *et al.* published a follow-up article to "The Overestimated Workweek? What

<sup>37</sup> Phillip M. Clarke, Cenzil G. Fiebig and Ulf-G. Gerdtham, "Optimal Recall Length in Survey Design," *Journal of Health Economics* 27 (2008), p. 1275 -1284.

<sup>38</sup> 135/(135+386)

<sup>39</sup> John P. Robinson and Ann Bostrom, "The Overestimated Workweek? What Time Diary Measures Suggest," *Monthly Labor Review*, August 1994, pp. 11-23.

<sup>40</sup> *Id.* at Table 2.



Time Diary Measures Suggest,” in 2011 based on 2003 through 2007 data.<sup>41</sup> This newer study confirmed that survey respondents systematically overstated their hours worked. A relevant feature of both the original Robinson and Bostrom article and the follow-up by Robinson *et al.* is their discussion of other research documenting survey respondents overstating the frequency or duration of other activities besides work. One study monitored households’ TV sets and found that TV ratings surveys overstated actual TV time by 20 to 50%. Another study surveyed swim and health club members on their club usage over the previous 12 months and compared the results to club records. Almost half of the respondents overstated their usage by 100% or more. A 1998-2001 diary study found that survey respondents overstated the time they spent on housework. Men overstated the time by over 100% (23 hours per week according to surveys versus 10 according to daily diaries) and women by nearly 100% (32 hours per week according to surveys and 17 according to daily diaries).

100. The most recent published research I have found demonstrating that survey responses may lead to extreme overstatement of the frequency of activities was conducted by researchers affiliated with the Federal Reserve Bank of Boston, “Measuring Household Spending and Payment Habits: The Role of ‘Typical’ and ‘Specific’ Time Frames in Survey Questions.”<sup>42</sup> In this article, the authors conducted a survey regarding respondents’ use of different payment mechanisms: debit cards, cash, credit cards or checks. Each respondent was surveyed four times at roughly three month intervals over the course of a year. Twice he or she was asked about payment activity over specific periods (a specific day within the past week, the past week in total, the past month in total and the past year in total) and twice he or she is asked about payment activity in typical periods. For example, respondents were asked how often they used cash in the typical day, typical week, typical month and typical year. Similarly, they were asked how much cash they spent in a typical day, typical week, typical month and typical year. If recall is accurate, survey responses should be consistent about the rate of use of a payment mechanism regardless of whether the question is framed in terms of a typical day, a typical week, a typical month or a typical year. The amount for a typical year should be 365 times the amount for a

<sup>41</sup> John Robinson, Steven Martin, Ignace Glorieux and Joeri Minnen, “The Overestimated Workweek Revisited,” *Monthly Labor Review*, June 2011, 43-53.

<sup>42</sup> Marco Angrisani, Arie Kapteyn and Scott Schuh, “Measuring Household Spending and Payment Habits: The Role of ‘Typical’ and ‘Specific’ Time Frames in Survey Questions,” Federal Reserve Bank of Boston, Working Paper Series No. 12-7, June 2012, available at <http://www.bostonfed.org/economic/wp/wpchron.htm>.



typical day, 52 times the amount for a typical week, and twelve times the amount for a typical month. The results demonstrated that responses were not consistent.

101. In Table 7 of their paper, the authors report the average responses regarding typical daily, weekly, monthly and annual use of the four different payment mechanisms. They report the results on an annualized basis, e.g., the average daily use of cash is multiplied by 365 so that it is comparable to the yearly cash usage number; the weekly average is multiplied by 52, etc. The table demonstrates that survey respondents' recall of the frequency at which they use debit cards, cash, credit cards or checks varies significantly depending on how the question is asked. When asked about frequency in terms of usage per day, respondents reported using the different payment mechanisms 19% to 119% more frequently than when the questions were framed in terms of frequency per year. Some or all of the survey averages must be inaccurate measures of the true frequencies at which the respondents used the various payment mechanisms. As the average daily answers and the average annual answers for a given group of respondents and a given payment mechanism are always inconsistent with each other, one or both must be wrong, and the magnitude of error is substantial.

102. Answers were also inconsistent when comparing a) answers to questions framed in terms of weeks to answers to questions framed in terms of years, b) weeks to months and c) months to years. The inconsistencies were not uniformly as striking as the inconsistencies between daily rates and annual rates, but they were often large and often statistically significant. For credit cards, the differences in annualized usage rates across time frames were always statistically significant. There were statistically significant differences in annualized usage rates for each of the four payment mechanisms when respondents reported typical usage by the day as opposed to the week, month or year. There were statistically significant differences in annualized usage rates when comparing the other time periods too, e.g., comparing annualized usage based on answers to the typical week to answers based on the typical month. However, the differences based on these other time comparisons were not consistently statistically significant across payment mechanisms.

## **B. RECALL WINDOW**

103. The extensive literature concerning recall bias eliminates any confidence that respondents' answers to Plaintiffs' survey accurately reflect the frequency at which they actually

engaged in the activities at issue in this litigation. Empirical literature spanning at least the last sixty years demonstrates that survey respondents systematically misstate the frequency at which events, activities and behaviors actually occurred. The magnitude of the error is not trivial either. As discussed above, various analyses spanning many different activities document errors exceeding 100%. This is systematic error, not random error, so the problem is not addressed by taking ever larger samples. For example, the Federal Reserve Bank study was based on a sample of 3,285, and all of the empirical literature I cite regarding recall bias involves samples much larger than 40.

104. Much of the literature regarding recall bias focuses on the length of the recall window, and the effect that that has on the magnitude of recall error. The recall window is the length of time over which the respondent is asked to recall events occurring. The academic community takes it as a given that a longer recall window implies greater recall error, and empirical research supports this idea.<sup>43</sup> This literature is relevant to the case at hand for two reasons. First, it tends to further establish that recall error is the norm for surveys that ask respondents to report on the frequency of past events. Second, the recall window in this case is exceptionally long. It is over eleven years for a majority of the respondents and several years for everyone else. In contrast, analysts concerned about the effect of the length of the recall window on the reliability of survey data have raised concerns about recall windows as short as three days.<sup>44</sup>

105. Many studies related to recall bias and optimal recall windows are in the health field. These studies may be particularly relevant because they often occur in the context of clinical trials related to serious illnesses. Respondents are motivated to report accurately. The fact that these studies routinely uncover recall bias that increases as the recall window widens shows that

---

<sup>43</sup> Phillip M. Clarke, Cenil G. Fiebig and Ulf-G. Gerdtham, "Optimal Recall Length in Survey Design," *Journal of Health Economics* 27 (2008), p. 1275, "It has been widely recognized that there is an inverse relationship between the length of time over which subjects are asked to recall prior use and the accuracy of the reported estimates. The longer the period of recall the greater the likelihood of error."

N.A. Connelly and T.L. Brown, "Effect of Recall Period on Annual Freshwater Fishing Effort Estimates in New York, Fisheries Management and Ecology, 18 (2011), p. 83, "Recall bias can influence estimates of fishing effort, with a general trend toward overestimation as the recall period increases."

<sup>44</sup> Daniel Feikin, Allan Audi, Beatrice Olack, Godfrey Bigogo, Christina Polyak, Heather Burke, John Williamson and Robert Breiman, "Evaluation of the Optimal Recall Period for Disease Symptoms in Home-Based Morbidity Surveillance in Rural and Urban Kenya," *International Journal of Epidemiology*, 39 (2010), p. 450: "A 2-week recall period underestimates true disease rates and health-care utilization. Shorter recall periods of 3 days in children and 4 days in adults would yield more accurate data."

recall bias is not due to respondents' failure to make their best efforts to answer accurately.

106. One example of a health-related study is Bennett, Amtmann, Diehr, and Patrick, "Comparison of 7-Day Recall and Daily Diary Reports of COPD Symptoms and Impacts."<sup>45</sup> In this study, the authors note that "the accuracy of short term retrospective recall (1-4 weeks) in measuring daily experiences has been tested in several areas." They cite ten studies in the health field and discuss the results of these studies. Correlation between data from daily diaries and recollections covering longer time periods varied between 0.33 to 0.89. Correlation is a measure of the degree to which data series move together and it varies from negative one to positive one. If data from daily diaries had been the same as data that respondents recalled over longer time spans, then correlation would have been equal to positive one. Consequently, in all of the studies that Bennett *et al.*, summarized there was significant recall error.

107. In their own study, Bennett, *et al.*, compared research participants' weekly review of the severity of their chronic obstructive pulmonary disease ("COPD") symptoms with data that the participants recorded twice a day with an electronic diary. They found that on average the patients' recollections of the severity of their breathing problems in the preceding week was 52% higher than what patients actually had experienced based on daily data. They found that on average the patients' recollections of activity limitations in the preceding week was 24% higher than what patients had actually experienced based on daily data. They found that on average patients' reports of feeling upset during breathing problems was 35% higher than what patients actually had experienced based on daily data.<sup>46</sup> Large errors may result when the recall window is as short as seven days.

108. Another study that is particularly relevant to this litigation relates to steel workers' survey responses regarding their annual earnings, hourly wage and hours worked.<sup>47</sup> Respondents were surveyed in 1983 about their pay and hours in 1981 and 1982. The study compared the survey

---

<sup>45</sup> Antonia Bennett, Dagmar Amtmann, Paula Diehr, and Donald Patrick, "Comparison of 7-Day Recall and Daily Diary Reports of COPD Symptoms and Impacts," *Value in Health* 15 (2012), 466-474.

<sup>46</sup> The comparisons in this paragraph are all derived from Table 6 of the Bennett, *et al.*, study. The authors ultimately conclude that weekly diaries may be useful for healthcare research, but that is because the weekly data can discern trends and changes in symptoms and thereby detect whether treatments are effective even though the weekly data are affected by significant recall error.

<sup>47</sup> G. J. Duncan and D. H. Hill, "An Investigation of the Extent and Consequences of Measurement Error in Labor Economics Survey Data," *Journal of Labor Economics* 3 (4) (October 1985).

responses to actual records. On average workers understated their annual earnings and hourly wages and overstated their hours worked. The errors were greater for 1981 than for 1982 meaning that a longer recall window introduced more error.

### C. DISTORTED PERCEPTIONS OF TIME

109. Plaintiffs' survey is not only unreliable because it is baseless to assume that respondents' recollections of the frequencies of past events reliably measure the actual frequencies of those events, but it is also unreliable because people's perceptions of time tend to be inaccurate too. Actual time and perceived time are two different things.<sup>48</sup> Moreover, people tend to overestimate wait time in particular, and wait time is a key element of Plaintiffs' claims.

110. Hornik published a study in the *Journal of Consumer Research* regarding customers' perceptions of the time spent waiting in line.<sup>49</sup> Researchers observed every n<sup>th</sup> shopper to enter a line at two branches of a chain supermarket. They timed the shoppers' actual wait time and also intercepted them after checkout and asked them to estimate how long they had waited. On average, respondents overestimated their wait-time by 36%. Depending on the type of line (multi-server, snake, express) only 3 to 5 percent of respondents measured their wait time accurately. Also depending on the type of line, 64% to 70% of respondents overestimated their wait time. Hornik also estimated a linear relationship between actual wait time and perceived wait time and found that perceived wait time grew 16% faster than actual wait time.

111. Antonides, Verhoef and Aalst conducted a study related to perceived telephone wait time.<sup>50</sup> Volunteers were selected from a shopping area and asked to call a phone number and request an information brochure from a financial institution. They were promised approximately \$2 for participating in the experiment. Participants were assigned either an 800 number or a 900 number to call. The callers dialing the 900 number would have a per-minute charge for the duration of the call deducted from their reward. The researchers varied the length of time that

<sup>48</sup> See Lorraine Allan, "The Perception of Time," *Perception and Psychophysics* 26 (1979), 340-354. Allan summarizes the time perception literature spanning the period 1930 through 1979. According to Allan at page 351, "There are considerable data that are consistent with a linear relationship between perceived time and stimulus time." In other words perceived time and actual time are related, but they are not the same thing.

<sup>49</sup> Jacob Hornik, "Subjective vs. Objective Time Measures: A Note on the Perception of Time in Consumer Behavior," *Journal of Consumer Research* 11 (June 1984), pp. 615-618.

<sup>50</sup> Germit Antonides, Peter Verhoef and Marcel van Aalst, "Consumer Perception and Evaluation of Waiting Time: A Field Experiment," *Journal of Consumer Psychology* 12 (3), pp. 193 – 202.

different participants were kept waiting—10 seconds, 20 seconds, 30 seconds, 1 minute, 2 minutes and 3 minutes. After the call was complete, the researchers administered a questionnaire asking among other things how long the participants believed they were on hold. Regardless of actual wait time, respondents' estimates were systematically overstated. Ten second waits were overestimated the most in percentage terms, on average 148%. Three minute waits were overestimated the most in absolute terms, on average 50 seconds or 28%. Generally speaking, persons who dialed a 900 number, and thus paid for the wait in the form of a lower reward, overestimated by more than the people who dialed an 800 number did. These results are particularly relevant to the case at hand because Wal-Mart drivers earn more if they are driving. Class members lose financially the longer they wait, whether they are earning unscheduled time or not, because it keeps them from driving.<sup>51</sup> This may cause them to further overstate time waiting or on other activities that keep them from driving.

112. Plaintiffs' survey presupposes that the respondents' recollections of how long activities took are a reliable measure of how long they actually took. There is no basis for this assumption. Experiments demonstrate that actual time and perceived time are different. Empirical research finds that people overestimate waiting time in particular and that the overestimate is magnified when respondents perceive that the wait costs them money such as in the case at hand where respondents tended to perceive that they could earn more driving than earning activity pay.

## **VI. FURTHER FLAWS AND ERRORS IN PLAINTIFFS' SURVEY**

113. Plaintiffs' survey is afflicted with other flaws or errors in addition to the ones discussed so far. The survey appears to be affected by nonresponse bias, and there is reason to be concerned that it is affected by social desirability bias and demand artifacts.

### **A. NONRESPONSE BIAS**

114. Nonresponse bias relates to the fact that the composition of the sample was not arrived at by an entirely random process. Not all of the class members whom the Plaintiffs subpoenaed at random were surveyed and deposed. In fact, the Plaintiffs had to attempt service on 104 class members in order to get 40 class members to respond. Six of the Plaintiffs' randomly selected class members were deceased. Two of the randomly selected class members reported that they

<sup>51</sup> Class members testified that this was indeed their perception. See, e.g., the Depositions of Stanley McCulley 04/28/16, pp. 28:17-29:10, and Dana Brittan 02/22/16, pp. 100:16-101:2.

were unable to appear due to health reasons. One of the randomly selected class member responded to his subpoena by opting out. Five of the randomly selected class members failed to appear for the survey notwithstanding that they had been subpoenaed, and Plaintiffs were unable to serve subpoenas on the remaining 50 of the randomly selected class members.

115. The 104 class members whom Plaintiffs sought to subpoena constitute a random sample of the class. The 40 class members who were surveyed and deposed do not constitute a random sample of the class unless the division of the 104 person random sample into respondents and non-respondents was also the result of a random process. For any given survey question, the difference between the average responses that would have been generated if all 104 had been surveyed and the average answers based on the 40 class members who were actually surveyed represents nonresponse bias. It is an unmeasurable gap between a reliable estimate of how the class would have answered the survey question (as represented by the average answer if all 104 of the randomly selected class members had answered) and the actual sample average. Although the magnitude of nonresponse bias may theoretically be small, there is no valid statistical basis to assume that that is the case.

116. Survey guidelines and reference manuals instruct researchers to conduct follow-up analyses to demonstrate that survey results are reliable when they reflect high degrees of nonresponse. For example, in the chapter entitled Reference Guide on Survey Research, the Federal Judicial Center, National Research Council's, *Reference Manual on Scientific Evidence*, Third Edition, states: "It is incumbent on the expert presenting the survey results to analyze the level of and sources of nonresponse, and to assess how the nonresponse is likely to have affected the results."<sup>52</sup> Previous versions of the Reference Guide on Survey Research cited guidelines from the former U.S. Office of Statistical Standards, stating that response rates below 50% should be regarded with significant caution as a basis for precise quantitative statements about the population from which the sample was drawn.<sup>53</sup> Similarly, the U.S. Office of Management and Budget Standards and Guidelines for Statistical Surveys ("OMB Guidelines") require that agencies "must appropriately measure, adjust for, report and analyze unit and item nonresponse

---

<sup>52</sup> Page 383.

<sup>53</sup> Diamond, S.S. (2000) "Reference Guide on Survey Research," in *Reference Manual on Scientific Evidence*, 2nd ed., Washington, D.C.: Federal Judicial Center, p. 239.

to assess their effects on data quality to inform users.”<sup>54</sup> The OMB Guidelines require federal agencies sponsoring surveys to conduct a nonresponse bias analysis if unit response rate is below 80% or if the item response rate is below 70% for any item used in the report.<sup>55</sup> The unit response rate for Plaintiffs’ survey was 38.5%. As unit nonresponse was below 70%, item nonresponse was below 70% for each question on the survey.

117. Although it is possible for nonresponse to be benign if there are no systematic differences between non-responders and responders, that does not appear to be the case here. Respondents were disproportionately former Wal-Mart employees (including as former employees those who are on extended leaves of absence). As of November 28, 2014, the final date for which comprehensive class-wide data on terminations were available in the discovery materials available to me, twenty-nine percent of the class members were former employees (235/806) while 40% of the survey respondents were former employees as of that time (16/40).<sup>56</sup> There is only a 9% chance that this difference in representation of former employees occurred randomly rather than being caused by systematic differences between former employees and current employees in their likelihood of responding to the subpoena.<sup>57</sup> This over-sampling of former employees has apparently biased the survey results towards longer frequencies and durations for the activities that Plaintiffs’ survey asks about.

118. Plaintiffs’ survey asks 24 questions about how frequently activities occurred or how long they lasted. I have separately tabulated average survey responses for current employees and former employees to these 24 questions. I show these tabulations in Figure 29. If former Wal-Mart employees were the same on average as current employees in terms of their perceptions of the frequency and duration of the various activities at issue in this case, then there should be a 50% chance for any given frequency or duration question on Plaintiffs’ survey that the average answer from former employees would be higher than the average answer from current employees. Given 24 frequency and duration questions, it would be expected that former employees’ average answers would be higher than current employees’ average answers 12 times. In fact, former

<sup>54</sup> Standard 3.2 accessed May 9, 2016 at

[https://www.whitehouse.gov/sites/default/files/omb/inforeg/statpolicy/standards\\_stat\\_surveys.pdf](https://www.whitehouse.gov/sites/default/files/omb/inforeg/statpolicy/standards_stat_surveys.pdf).

<sup>55</sup> OMB Guidelines 1.34 and 1.35.

<sup>56</sup> WMBryan-00029436.

<sup>57</sup> This 9% estimate is based on the binomial probability of selecting 16 or more former employees in 40 trials if the true likelihood of selecting a former employee is 235/806.



employees' average answers about frequency or duration were higher 19 times, *i.e.*, 79% of the time rather than 50% of the time. The probability of former employees reporting higher frequencies or durations 19 times or more when 24 questions were asked is approximately 0.3%.<sup>58</sup>

**Figure 29**  
**Survey Responses by Employment Status**

Question	Former Employees		Current Employees		Difference of Mean Responses (Former-Current)	Percentage Difference of Means
	Mean	Number of Responses	Mean	Number of Responses		
3a Usual pretrip length	15.00	16	13.35	23	1.65	12%
4a Usual post-trip length	13.13	16	10.00	23	3.13	31%
5a # roadside/weigh inspections in usual month	2.55	13	1.54	12	1.01	65%
5b Usual inspection length	22.67	15	20.53	18	2.14	10%
6a # washings in 10 usual trips	3.33	15	2.20	15	1.13	52%
6b Usual washing length	36.25	16	26.33	21	9.92 *	38%
7a # weighings outside Wal-Mart in 10 usual trips	2.07	14	2.04	13	0.03	2%
7b Usual length of weighing outside Wal-Mart	23.75	16	21.25	22	2.50	12%
8a # adjustments outside Wal-Mart in 10 usual trips	1.88	12	1.13	15	0.74	65%
8b Usual time to adjust	32.17	15	28.61	18	3.56	12%
9a # refuelings at Wal-Mart in 10 usual trips	6.33	15	5.02	21	1.31	26%
9b Usual time to refuel at Wal-Mart	18.67	15	15.28	23	3.38	22%
10a # refuelings outside Wal-Mart in 10 usual trips	1.54	14	2.03	17	-0.49	-24%
10b Usual time to refuel outside Wal-Mart	20.31	16	21.82	22	-1.51	-7%
11a # meetings at start of day or trip in 10 usual trips	8.47	15	8.10	21	0.37	5%
11b Usual length of meeting at start of day or trip	17.33	15	6.67	21	10.67 **	160%
12a # meetings at end of trip in 10 usual trips	6.33	15	6.50	20	-0.17	-3%
12b Usual length of meeting at end of trip	12.33	15	5.98	21	6.36 *	106%
13a # times waiting for loading in 10 usual trips	3.46	13	2.41	16	1.06	44%
13b Usual time waiting for loading	62.88	13	80.16	16	-17.27	-22%
14a # times waiting for unloading in 10 usual trips	3.35	13	2.50	16	0.85	34%
14b Usual time waiting for unloading	38.57	14	43.55	19	-4.98	-11%
15a # rest breaks in 10 usual trips	10.37	15	9.30	20	1.07	11%
15b Usual rest break length	13.17	15	12.19	21	0.98	8%

Source: 2016 Plaintiff Questionnaires; Depositions of Subpoenaed Drivers

**Notes:**

1. The means reported are the averages of survey responses for which a mean could be computed, *i.e.*, responses with a single number or a range of numbers.
2. \* denotes significance at the 10% level, \*\* denotes significance at the 5% level, and \*\*\* denotes significance at the 1% level.
3. The number of responses to a given question may be less than the number of current and former employees to the extent that some drivers did not respond, or provided a non-numeric response, to that question.
4. If an answer is vague with a number, such as "60 depends on...", then we take the number as the answer.
5. If an answer is nonnumeric such as "varies", "do not recall", "unknown", etc., we treat it as a blank.
6. If an answer is in the format of ">#" or "#+", we take # as the minimum, and blank as the maximum.
7. If an answer is in the format of "<1", we take 1 as the maximum, and 0 as the minimum.
8. If the answer does not match the question, we treat it as a blank.
9. If an answer is "no", we assume the numeric response is 0. If an answer is "yes" and the numeric response is blank, we assume the response is blank. If an answer is "yes" and the numeric response is 0, we assume the response is 0.
10. Based on deposition testimony, some drivers' survey responses were updated. Kenneth Nevarez answered "no" for some questions and then provided numeric answers. We use those numbers since he stated in his deposition that he intended to answer "yes". John Rivero answered "no" for 11 and 12, and then provided numeric answers. We take "no" as answers and treat the numbers as zeros since according to his deposition, he intended to answer "no".

119. The differences between the frequencies and durations that former employees reported on average and the frequencies and durations that current employees reported on average were not merely significant in the statistical sense, *i.e.*, it is not merely the case that they are not attributable to chance, but they were also significant in the practical sense. The differences correspond to large differences in damages estimates depending on whether the groups ought to be analyzed separately or one or both groups' answers are invalid. In Figure 29 I show mean survey responses

<sup>58</sup> Based on the binomial distribution where each survey question is characterized as an independent draw with a 50% probability that current employees will report a higher frequency or duration and a 50% chance that former employees will report a higher frequency or duration.



for current employees and former employees regarding those activities which former employees said occurred more frequently or lasted longer than current employees said. I also show the percentage difference between the frequency or duration reported by former employees and the frequency or duration reported by current employees. An “x” percent overstatement in the average frequency of a task would imply an “x” percent overstatement in damages attributable to that task under Plaintiffs’ damages theory. Similarly, an “x” percent overstatement in the average duration of a task would also imply an “x” percent overstatement in damages attributable to that task under Plaintiffs’ damages theory. Overstating both frequency and duration would have a compound effect. Consequently, Figure 29 indicates that damages would be much different depending on whether they were based on current employees’ responses, former employees’ responses or all responses.

120. As shown in Figure 29, former employees reported activities taking up to 160% more time than current employees said they took (initial workday meetings with driver coordinators). Similarly, former employees reported activities occurring up to 65% more frequently than current employees said they occurred (CHP/DOT roadside/weigh inspections and adjustments when weighing outside of Wal-Mart). Because differences in frequencies and durations would compound, damage calculations under Plaintiffs’ theory related to initial workday meetings with driver coordinators would be 172% higher if based on former drivers’ recollections rather than on current drivers’ recollections.

121. To summarize, former employees are overrepresented in the sample. Former employees tend to say that the activities that the Plaintiffs surveyed about occurred more frequently and lasted longer than current employees say. The difference between former employees and current employees in reported frequencies and durations of these activities is statistically significant. Consequently, Plaintiffs’ survey responses are biased towards higher reported frequencies and longer durations of the activities at issue due to nonresponse bias. Current employees were less likely than former employees to appear at deposition in response to the subpoenas. Current employees who did appear were likely to report that the activities at issue occurred less frequently and took less time than former employees would say.

122. The differences between former employees’ survey responses and current employees’ survey responses not only raise the issue of nonresponse bias, but they also raise questions about

whether damages can be calculated accurately for former employees and current employees as one single group and whether the former employees' responses are accurate at all. The only explanations for the different responses are either that former employees and current employees actually had meaningfully different experiences, in which case damages should be calculated for the groups separately, or that one or both groups of respondents' answers are inaccurate.

## **B. SOCIAL DESIRABILITY BIAS**

123. Social desirability bias refers to the tendency for survey respondents to underreport engaging in socially undesirable conduct (*e.g.*, illicit drug use) and over-report engaging in socially desirable conduct (*e.g.*, voting).<sup>59</sup> Researchers have demonstrated social desirability bias by comparing survey responses to objective data regarding the incidence of socially desirable or undesirable behavior. For example, in a study of the 1976 presidential and 1978 off-year elections, 72% of survey respondents said they voted in 1976, while voting records showed that only 61% actually had done so.<sup>60</sup> Similarly, 55% said they had voted in 1978 while only 43% had done so.<sup>61</sup> Another study found that 39% over-reporting of having voted in a primary, 19% over-reporting of having a library card and 15% over-reporting of having registered to vote. The same study found 32% underreporting of having filed for bankruptcy and 47% underreporting of having been charged with drunk driving.<sup>62</sup>

124. Social desirability bias is relevant to Plaintiffs' survey because some of the activities asked about are related to good citizenship, perceived regulations and social consciousness, specifically safety inspections and truck washing. There may be over-reporting of the frequency and duration of truck washing and safety inspections from Plaintiffs' survey due to social desirability bias.

## **C. DEMAND ARTIFACTS**

125. Demand artifacts are biases in study participants' behavior related to their awareness of the purposes of the study. An obvious demand artifact that may bias Plaintiffs' survey is self-interest. Respondents may have known that greater reported frequencies and greater reported durations would support higher damages. In fact, Plaintiffs had sent several of the randomly

<sup>59</sup> Groves et al. *supra*, p. 168.

<sup>60</sup> Robert Groves, *Survey Errors and Survey Costs*, 2004, 438-9.

<sup>61</sup> *Ibid.*

<sup>62</sup> *Idem*, p. 440.

selected class members another survey that said on its face that its purpose was to assist in assessing damages.<sup>63</sup> Self-interest is not the only potential demand artifact that could bias the survey results. Some respondents may have consciously or unconsciously biased their answers upwards either to avoid letting other class members down or in the process of over-correcting for perceived memory problems. It is not always predictable what the magnitude or sign of the bias associated with demand effects will be. That is why manuals and treatises discussing best-practices in survey methodology discuss various protocols to minimize demand effects, the first of which is to avoid informing the respondent about the purpose of the survey or the identity of the sponsor.<sup>64</sup>

## VII. SUMMARY OF FLAWS RELATED TO PLAINTIFFS' SURVEY DATA

126. Plaintiffs' survey is irreparably flawed for purposes of proving injury and quantifying damages for at least the following eight reasons.

127. First, Plaintiffs' survey fails to adequately address certain topics that must be answered to prove the fact of injury or to quantify damages. These include concurrent activities, the frequency and duration of paid rest breaks and the amount of layover time class members spent outside of Wal-Mart's control such as at hotels, restaurants, movies and other entertainment venues or at home. I understand further that Wal-Mart's position is that time during which neither the class member nor Wal-Mart perceived that the class member was under Wal-Mart's control must also be isolated and removed from that class member's compensable layover time. Plaintiffs' survey does not address this topic either notwithstanding deposition testimony from randomly selected class members that they chose to sleep in their trucks for their own personal

---

<sup>63</sup> Letter dated April 3, 2014 from Wagner, Jones, Kopfman & Artenian, LLP addressed to "Putative Class Member / Wal-Mart Driver." Three randomly selected class members testified to receiving and filling out the survey. See Exhibit 3 to the Deposition of Todd Brown, 03/15/16; Exhibit 3 to the Deposition of Charles McLaughlin, 03/18/16; Exhibit 1 to the Deposition of Kevin Putnam, 03/23/16. Another randomly selected class member, Michael Rand, testified that he received and thought that he filled out an earlier survey (Deposition of Michael Rand, 02/22/16, pp. 37:21-38:7).

<sup>64</sup> See, e.g., Shari Seidman Diamond, "Reference Guide on Survey Research," in *Reference Manual on Scientific Evidence, Third Edition* (2011), pp. 410-411: "To ensure objectivity in the administration of the survey, it is standard interview practice in surveys conducted for litigation to do double-blind research whenever possible: Both the interviewer and the respondent are blind to the sponsor of the survey and its purpose. Thus, the survey instrument should provide no explicit or implicit clues about the sponsorship of the survey or the expected responses. Explicit clues could include a sponsor's letterhead appearing on the survey..."

reasons rather than perceiving that Wal-Mart required it<sup>65</sup> or that they would need Wal-Mart's permission to spend their layover time elsewhere.

128. Second, Plaintiffs' survey was confusing. The questions were vague and ambiguous. The survey responses and follow-on deposition testimony show that not all respondents were interpreting questions in the same way. The responses demonstrate that the respondents were confused by the questions. For example, some respondents' answers were mutually inconsistent.

129. Third, respondents admitted that they were unable to answer the survey questions reliably. The experiences that they were being asked to average and summarize varied too much from day to day and year to year or occurred too far in the past for the respondents to consider their answers to be reliable.<sup>66</sup> Frequently, respondents refused to give specific numeric answers, answering with ranges or qualitative responses instead. Respondents occasionally changed their answers at deposition. Some respondents explicitly labeled their own answers as unreliable. A few respondents had answered an earlier survey from Plaintiffs' counsel that had some comparable questions.<sup>67</sup> These people rarely answered any of the comparable questions the same way twice.

130. Fourth, responses to the Plaintiffs' survey indicate that respondents' experiences were dissimilar from one another. Assuming *arguendo* that the respondents' average answers about the frequencies and durations of various activities were indicative of the corresponding averages at which class members actually engaged in those activities and how long on average class members spent on those activities, it would be still be highly inaccurate to rely on the survey responses to infer injury and calculate damages. Too many class members were too far from

<sup>65</sup> Depositions of James Batham 05/17/16, p. 118:4-18; David Evans 03/22/16, p. 22:4-9; Robby Robinson 05/10/16, p. 58:14-25.

<sup>66</sup> See, e.g., Deposition of Bobby Carr, 04/12/16, p. 104:10-14: "Q. · And you can't say from week-to-week or month-to-month? A. · Day-to-day. Q. · Day-to-day, it all changes? A. · Yes." Also see Deposition of Robert Benavidez, 03/15/16, p. 123:14-18: "Q. · Okay. · Just one moment. · When you fill out your survey as we discussed you noted that various tasks varied too much and you couldn't provide an estimate; is that right? A. · Yes." Several survey respondents testified that they would not want to be tied to the estimates that they provided on the survey under penalty of perjury. See, e.g., Deposition of Mark Alumbaugh, 03/18/16, pp. 53:21-54:8; Deposition of Charles McLaughlin, 03/18/16, 36:18-37:5; and Deposition of Kevin Putnam, 03/23/16, p. 56:10-20. Drivers also testified that their memory has faded, affecting their ability to reliably recall the frequency and duration of work performed throughout the entire class period. See, e.g., Deposition of Raymond Byer, 02/29/16, p. 96:18-22 and Deposition of Eric Nettles, 02/26/16, p. 109:2-24.

<sup>67</sup> Exhibit 3 to the Deposition of Todd Brown, 03/15/16; Exhibit 3 to the Deposition of Charles McLaughlin, 03/18/16; Exhibit 1 to the Deposition of Kevin Putnam, 03/23/16.

average for the average to generate an accurate measure of class damages.

131. Fifth, Plaintiffs' survey is afflicted by a significant degree of sampling error. Class members' experiences are so varied that a sample size of 40 leaves a tremendous amount of sampling error remaining. This means that for a given question the average answer that would result if the entire class were surveyed may be meaningfully different than the average answer for the sample of respondents. Not only is the class average an unreliable proxy for most individuals' experiences due to individuals being so different from one another, but the survey average does not give a precise measure of what the class average is.

132. Sixth, Plaintiffs' survey attempts to measure the frequency at which class members engaged in activities in the past and the duration of those activities by asking the survey respondents about their experiences, but Plaintiffs offer no evidence that surveys are a reliable method to determine the actual frequencies and durations at which survey respondents engaged in past activities. To the contrary, there is a large academic literature indicating that people's survey responses are an unreliable measure of the frequency and duration of past activities. People routinely misstate the frequency at which past activities occurred. These errors are not random, so they do not average out as the sample size increases. Nor are the errors trivial. Average answers from surveys may be off by hundreds of percent. The errors are not due to a lack of effort on respondents' part either. It does not matter that respondents answer to the best of their ability. Even when their lives are literally at stake, survey respondents answer surveys about past frequencies wrongly. Moreover, the error commences when they are asked to recall events as recently as three or four days in the past. Surveys about events in the more distant past than that are even less reliable. Plaintiffs' survey asks about events that occurred over eleven years before the survey date.

133. Seventh, people not only misremember how frequently they engaged in prior activities, but they also systematically misperceive how long past events took. They tend to overestimate time spent on passive activities in particular.

134. Eighth, Plaintiffs' survey is sensitive to other biases discussed at length in survey literature in addition to survey respondents' general inability to recall the frequency or duration of prior experiences. These biases include but are not necessarily limited to nonresponse bias, social desirability bias and demand artifacts. Plaintiffs' survey data are only reliable if the survey

is unaffected by them, and I am unaware of any sound basis to assume that these biases are absent. To the contrary, there is empirical evidence that Plaintiffs' estimated frequencies and durations are overstated due to nonresponse bias because of overrepresentation of former Wal-Mart Private Fleet Drivers in the survey sample. There are conceptual reasons to expect that they are afflicted by social desirability bias and demand effects as well.

### **VIII. DR. PHILLIPS' DAMAGES AND PENALTY ANALYSES**

135. Dr. Phillips estimated damages and penalties on behalf of Plaintiffs.<sup>68</sup> Dr. Phillips' damages estimates ignore a) that class members were admittedly relieved of all work duties during some unmeasured fraction of their layovers so that it is not accurate to treat 10 hours of layover time as 10 hours of time that should have been paid at minimum wage, b) that class members actually did take paid rest breaks so it is not accurate to simply assume that all days worked must be days in which rest breaks were missing, and c) that various purportedly unpaid activities occurred concurrently with each other and with paid activities. Ignoring these facts renders all of Dr. Phillips' damages analyses and penalty calculations inaccurate. Dr. Phillips' analyses are further flawed in other significant ways including but not limited to their statistically invalid extrapolations from very limited deposition testimony and their adoption of baseless assumptions regarding class members' experiences. I discuss each of Dr. Phillips' damages and penalty calculations and their major flaws in the order that they appear in Dr. Phillips' report.

#### **A. LAYOVERS**

136. Dr. Phillips assumed that "drivers were paid \$42 for each 10 hour layover but should have been paid minimum wage for that time."<sup>69</sup> Dr. Phillips identified class member layovers accounted for in the dispatch data that Wal-Mart produced in discovery. He calculated how much 10 hours' worth of pay would have been at the prevailing minimum wage as of the time of each layover, and he deducted \$42 from that to get a preliminary estimate of damages for each layover accounted for in the dispatch data. Based on Wal-Mart's dispatch data and the assumption that each layover should have earned drivers 10 hours of pay at the minimum wage, Dr. Phillips calculated \$25.6 million in layover-related damages.

<sup>68</sup> Mr. Edward Garcia signed an expert report discussing penalty calculations. Dr. Phillips included those calculations in his report, and included Mr. Garcia's report as Appendix C to his own report. I will refer to the penalty analysis as Dr. Phillips' although it appears to originate with Mr. Garcia.

<sup>69</sup> Phillips Report, p. 9.

137. Separately, Dr. Phillips had sampled paper Wal-Mart records. These paper records identified class members who were not included in the dispatch data produced in discovery. In addition, Dr. Phillips claims to have found more layovers in the sampled paper records than in the electronic data for class members who were represented in both. Dr. Phillips purported to have conducted an analysis demonstrating that his sample of paper records provided the more accurate count of layovers for the class members and pay periods appearing in the sample. Dr. Phillips proposed alternative adjustments to his \$25.6 million estimate to account for underreporting of layovers in the dispatch data attributable to the exclusion of class members from the dispatch data and the supposed underreporting of layovers taken by the class members who were included in the dispatch data. These adjustments range from an upward adjustment of 5.35% to account solely for class members missing from the dispatch data to an adjustment of 16.98% to account both for class members not being included in the dispatch data and also for the purported undercounting of layovers by class members who were in the dispatch data. After adjusting the dispatch-based layover estimate up by 5.35% to 16.98%, Mr. Phillips computes prejudgment interest at 10%.

138. Dr. Phillips' fundamental assumption is that all layover time is compensable. As discussed above, class members have testified that they took layovers away from their trucks with and without having given notice to Wal-Mart. Many class members also testified that they were not subject to any restrictions or control while in their trucks.<sup>70</sup> Class members also testified to having spent time away from their trucks engaged exclusively in personal activities such as eating at restaurants and socializing even on layovers when they ultimately slept in their trucks. If layover time spent away from Wal-Mart's trucks at home, at motels, friends' houses, at restaurants, at entertainment venues, exercising, and so on is not compensable, or if time spent in the trucks without restrictions or control is not compensable, then Dr. Phillips' layover analysis is invalid.

139. Wal-Mart witnesses and randomly selected class members have also testified that class members were free to spend their layovers outside of their trucks wherever they chose without needing any approval from Wal-Mart. Conversely, to my knowledge, no class members have said that they had wanted to spend time away from their trucks but that Wal-Mart refused to

---

<sup>70</sup> See, e.g., Deposition of Charles McLaughlin, 03/18/16, p. 44:4-10.



allow them to do so. If the Court determines that persons who did not perceive that they were under Wal-Mart's control could not actually have been under Wal-Mart's control during their layovers and thus were not injured, such a determination would be another reason that Dr. Phillips' layover calculations are invalid.

140. Leaving aside that not all layover time was time spent in Wal-Mart's trucks under Wal-Mart's control, at least one of Dr. Phillips' proposals to scale up of the layover damages estimates derived from the dispatch data purportedly to correct for its supposed deficiencies is methodologically invalid for additional reasons. Essentially, Dr. Phillips compared apples to oranges by failing to correctly account for the time span covered by his paper sample data and the dispatch data. Also, the correction to account for class members who are missing from the dispatch data is conceptually invalid. Most class members are in the dispatch data. A small fraction is not. Rather than estimate layover damages for the missing class members, Dr. Phillips adjusts up the damages estimates for the included class members. This was incorrect. The fact that data are missing to calculate one class member's damages accurately does not mean that some other class member's damages estimate is made more accurate by an upward adjustment. I explain both problems below.

141. Dr. Phillips sampled greenbars (paper summaries of Wal-Mart pay records) and created what I will call the Phillips Layover Database. His sample included data regarding 792 drivers in certain pay periods occurring between 5/1/2004 and 7/25/2015. The dispatch data which Wal-Mart produced in discovery covers 11/4/2004 through 1/12/2016, but it is missing data on certain class members including 86 class members who appear in the Phillips Layover Database. To determine that the Phillips Layover Database was reporting more layovers per driver than the dispatch data showed, Dr. Phillips should have compared the layovers that drivers in his sample accrued in his sampled pay periods between 11/4/2004 and 7/25/2015 to the number of layovers reported in the dispatch data for those same drivers for the same sampled pay periods. That is not what Dr. Phillips did to generate his largest proposed correction, a scale factor of 16.98%. To generate his 16.98% upward adjustment factor, Dr. Phillips compared layovers that drivers in the Phillips Layover Database accrued as early as 9/4/2004 to the layovers reported in the dispatch data. By including layovers that accrued before the dispatch data began, Dr. Phillips guaranteed that certain layovers that were in the Phillips Layover Database would be missing from the dispatch data so that it would appear



that the dispatch data was not recording all of the layovers for those drivers that it did cover.<sup>71</sup>

142. Dr. Phillips' correction to account for missing class members apparently compensates class members who are included in the dispatch data for harm allegedly suffered by class members who are not in the dispatch data. Dr. Phillips estimates that the class members who were in his sample but missing from the dispatch data increased the count of layovers in his sample by 5.35%. Consequently, he proposes to adjust upwards the damages estimates for each of the 712 persons who are represented in the dispatch data by 5.35%. In other words, Dr. Phillips estimates that 86 class members who are not accounted for in Wal-Mart's data would have damages equal to 5.35% of the amount that he estimates for 712 class members who are accounted for in Wal-Mart's data. However, Dr. Phillips does not describe any method of allocating this extra 5.35% of estimated layover damages among the additional 86 class members who supposedly deserve it. He apparently allocates it to the 712 class members included in the dispatch data instead. If Dr. Phillips' layover analysis were otherwise correct, this correction would apparently, purposely overcompensate the 712 class members who were included in the dispatch data by 5.35% each without mitigating at all the supposed under-compensation owed to the 86 class members outside of the dispatch data.

## **B. PRE-TRIP INSPECTIONS**

143. Dr. Phillips based his pre-trip inspection damages on the assumption that each class member performed a pre-trip inspection at the commencement of each work day. He assumed that class members should have been paid at the minimum wage for this time, but that they were not paid at all. Dr. Phillips identified class members' work days based on Wal-Mart's dispatch data. He calculated foregone pay based on various alternative assumptions about how long each pre-trip inspection took. For one set of estimates, he assumed that the inspection took eight minutes, purportedly because there is a Wal-Mart video of a pre-trip inspection and the inspection takes eight minutes in that video. In another set of estimates, Dr. Phillips assumed that inspections take 15 minutes because he was told that another expert would opine that that was how long pre-trip inspections would take.<sup>72</sup> For another set of estimates he purportedly used the

<sup>71</sup> Dr. Phillips' 16.98% figure also appears to be affected by an arithmetic error. Based on his workpapers, the figure ought to be 16.15% if the analysis were otherwise correct. See "16.9814% Adjustment Calculation).xlsx".

<sup>72</sup> See Report of V. Paul Herbert, pp. 11 – 12.

median pre-trip inspection figure from among class-member-deponents who were asked about pre-trip inspection duration (12.5 minutes according to Dr. Phillips). For a fourth estimate, he purportedly used the mean time estimate from among class-member-deponents who were asked about pre-trip inspection duration (14.4 minutes according to Dr. Phillips). For his final estimate, Dr. Phillips assumed that the inspection took one minute.<sup>73</sup>

144. As a threshold matter, Wal-Mart asserts that hook and depart pay is one of the types of pay that was intended to compensate drivers for the set of activities that occur at the commencement of a journey prior to mileage pay accruing. Certain class members testified that they also perceived that that was what hook and depart pay was for.<sup>74</sup> If the Court finds that in instances when Wal-Mart and class members both perceived that the pre-trip inspection was a subset of the activities that constitute a depart or hook and no further pay beyond “hook” or “depart” pay is required, then Dr. Phillips’ pre-trip damages estimates are invalid. Dr. Phillips’ damages estimates are not limited to persons who disbelieved that pre-trip inspections were not part of the activities for which they were compensated via “hook” or “depart” pay. In fact, he has not identified a single class member who testified to such a belief. Similarly, if the Court finds more broadly that pre-trip inspections are part of “hook” or “depart” for all class members regardless of what isolated, individual class members may have thought, then Dr. Phillips’ pre-trip damages estimates are also invalid under this condition.

145. Dr. Phillips’ various estimates of the typical duration of pre-trip inspections are all speculative. The video that he relies on does not establish that eight minutes is the typical length of time that class members actually spent. Similarly, V. Paul Herbert opines on Plaintiffs’ behalf that an inspection should take 15 minutes if done properly, but that does not mean that that is how long class members’ pre-trip inspections actually took in the field.

146. The average and median estimates of duration from class member testimony are all subject to the flaws discussed above wherein I discussed Plaintiffs’ survey. There is no basis to assume that such estimates are reliable given people’s tendency to misperceive the duration of activities; and there is no reason to assume that the responses are unaffected by social desirability

<sup>73</sup> “Deposition Responses.xlsx” and “Preliminary Loss Calculations by Sub-Category.xlsx.”

<sup>74</sup> Depositions of Steven Davis (03/22/16) pp. 163:17-164:2, Robert Mosley (03/08/16) pp. 116:11-21 and 174:20-176:1, Michael Baker (04/15/16) p. 53:6-13 and Anthony Fantasia (03/07/16) p. 43:9-15; and the Depositions of drivers Stanley McCulley (04/28/16) pp. 57:17-58:22 and Angel Vasquez (05/04/16) pp. 24:8-26:6.

bias or demand artifacts. Most importantly, the class members who happened to be asked at deposition about the duration of pre-trip inspections do not constitute a random sample. There is no statistically valid basis to assume that the mean or median response from the 27 deponents who were asked is representative of the time that the other 700+ class members spent on the activity. In particular, the mean and median response is affected by Plaintiffs' counsels' decisions about whether to ask about pre-trip inspection duration after having read the class members' survey responses, listened to the class members' responses to Wal-Mart's deposition questions and possibly after having conferred with the deponent beforehand.

147. Dr. Phillips ignores the evidence from his firm's own survey that length of pre-trip inspections varied significantly from class member to class member. Even if Dr. Phillips' estimate of the mean or median length of pre-trip inspections turned out to be correct, each class members' pre-trip inspection damages could be misstated markedly. For example, the minimum survey response regarding pre-trip inspection length was five minutes while the maximum was 30 minutes. Leaving aside Dr. Phillips' *de minimis* one minute calculation and taking the survey responses at face value, basing damages on the mean response from the larger sample of all survey respondents rather than the mean response among the subset of survey respondents who were also asked about pre-trip inspections in their depositions, damages calculations would misstate damages at the individual level by -73% (should be 30 minutes but compensated for eight) to +200% (should be five minutes but compensated for 15) for the survey respondents alone. The misstatements for the class as a whole would undoubtedly be even greater as it is implausible that Plaintiffs' happened to sample the class members who spent the most and least time on average doing pre-trip inspections.

### C. POST-TRIP INSPECTIONS

148. Dr. Phillips defined "post-trip" inspections to be safety inspections that class members conduct at the end of each workday. He offered three post-trip inspection damages estimates. One based on post-trip inspections taking 5 minutes on average, supposedly based on some other Plaintiffs' expert's testimony, another based on the median deposition answer among survey respondents who were asked about post-trip inspection length at deposition (12.5 minutes according to Dr. Phillips) and a third based on the mean deposition response from among these same class member/deponents (10.6 minutes according to Dr. Phillips).

149. Dr. Phillips assumed that post-trip inspections occurred at the end of each day that each class member worked. He provides no basis for this assumption, and it contradicts the discovery record. One class member testified explicitly that he did not perform a safety inspection at the end of each workday.<sup>75</sup> Plaintiffs' survey did not ask how frequently class members conducted post-trip inspections. Rather it asked whether they usually performed a post-trip inspection at the end of the workday. Given that the question merely asked whether respondents usually conducted post-trip inspections it is impossible to determine from Plaintiffs' survey how frequently post-trip inspections actually occurred. To my knowledge, there is no other source for information about the frequency of end of workday inspections available elsewhere in the discovery record either.

150. Wal-Mart and at least some class members also testified that post-trip inspections were part of the "arrive" activity.<sup>76</sup> According to them, class members were compensated for at least some end of workday post-trip inspections through arrive pay. Similar to the discussion above in the context of pre-trip inspections, Dr. Phillips' post-trip inspection calculations are invalid if the Court finds that class members who agreed with Wal-Mart that arrive pay compensated for post-trip inspections are not eligible for additional pay for post-trip inspections that were conducted as part of arrives. Dr. Phillips has not identified which class members held this belief or that any class members believed otherwise. Nor has he estimated the percentage of post-trip inspections that occurred outside of arrives.

151. Dr. Phillips' estimates of the typical time post-trip inspections took are all speculative. The five minute estimate is arbitrary and untethered to class members' actual experiences. The deposition-based estimates all suffer from the same shortcomings discussed above related to pre-trip inspections.

152. Similarly, Dr. Phillips overlooks the apparent variation in experiences across class members. Survey respondents who said they did usually conduct post-trip inspections testified to average durations of anywhere from 5 to 30 minutes as did deponents who were asked at

---

<sup>75</sup> Deposition of Charles Fox 02/17/16, p. 36:4-6.

<sup>76</sup> Deposition of Michael Baker 04/15/16, pp. 13:8-14 and 53:6-13; Deposition of Mark Alumbaugh 03/18/16, p. 79:9-15.

deposition.<sup>77</sup> Taken together, Dr. Phillip's demonstrably false assumption that safety inspections occurred at the end of each workday, his assumption that none of these inspections were compensated and his failure to account for the wide variation in the amount of time on average that safety inspections took all mean that his estimate of post-trip inspection damages are entirely unreliable at either the class level or individual level.

#### **D. PAID REST BREAKS**

153. Dr. Phillips proposed alternative methods of calculating damages related to purportedly unpaid rest breaks. None of his methods are reliable.

154. Dr. Phillips assumed that drivers never took paid rest breaks; therefore, rest break damages accrued each workday. This conflicts with the discovery record. As discussed above, class members testified that they took paid rest breaks while on unscheduled time, during live loads and during live unloads, among other examples. Class members also testified that they understood that they were allowed to take rest breaks at their personal discretion.<sup>78</sup> Under these circumstances, Dr. Phillips' rest break damages estimates are all inaccurate and unreliable. It is impossible to say for any given class member, how many of his or her workdays were lacking in paid rest breaks. Perhaps more importantly since workers may waive rest breaks, it is impossible to say for any given class member, how many of his or workdays, if any, were lacking in paid rest breaks.

155. Dr. Phillips proposed five rest break damages estimates. Based on my review of Dr. Phillips' workpapers, the first four are based on testimony from the non-random sample of class members who were asked at deposition how frequently they stopped driving to take unpaid rest breaks and how long such unpaid rest breaks lasted.<sup>79</sup> These four estimates differ primarily in whether Dr. Phillips based his calculations on mean responses or median responses and whether he based them on deponents whose testimony was framed in terms of frequency per 10 usual trips or on deponents whose testimony was framed in terms of frequency per unit of time, *e.g.*, number of unpaid rest breaks per day or week. These different assumptions yielded four different estimates of the frequency at which class members stopped for unpaid rest breaks. Dr. Phillips

---

<sup>77</sup> "Deposition Responses.xlsx."

<sup>78</sup> See Appendix 2.

<sup>79</sup> "Preliminary Loss Calculations by Sub-Category.xlsx" and "Deposition Responses.xlsx."

also calculated the mean and median duration of such breaks from the deposition testimony. Based on these estimated frequencies and durations and the minimum wage, Dr. Phillips generates four estimates of how much class members would have earned if they had been paid for these unpaid rest breaks at the minimum wage.

156. Dr. Phillips' workpapers indicate that 13 deponents testified regarding the frequency of unpaid rest breaks they took per 10 usual trips, 12 deponents testified regarding the frequency of unpaid rest breaks they took per unit of time, and 17 deponents testified regarding the duration of their unpaid rest breaks. Thus each of these rest break damages estimates is based on the deposition testimony of a very limited number of class members. These deposition-based estimates are subject to the same flaws discussed above in the context of pre-trip and post-trip inspections. They are subject to recall bias and demand artifacts. As none of these groups of class members constitutes a random sample, there is no statistically valid basis to extrapolate the mean or median responses to the class as a whole. To the contrary, as discussed above, whether Plaintiffs' counsel would choose to ask a deponent about rest break duration or rest break frequency could be influenced by the deponent's survey responses about frequency and duration of rest breaks and his responses to Wal-Mart's deposition questions.

157. These *unpaid* rest break analyses discussed in the preceding paragraphs have nothing to do with *paid* rest. Here, Wal-Mart *did* authorize and permit paid breaks according to class member testimony. Here, Plaintiffs' expert does not identify any evidence that any class member worked a single day without the opportunity and permission to take 10 minutes of paid rest for each four hours of work.

158. In addition to paid rest breaks, class members also took unpaid rest breaks at their discretion. The frequency and duration of a class member's unpaid rest breaks was entirely at the class members' discretion and tell nothing about the frequency and duration of his or her paid breaks. Consequently, Dr. Phillips' analysis of *unpaid* rest breaks has nothing to do with damages in this case. Dr. Phillips' analysis tells nothing about whether there is a shortfall in the number of *paid* rest breaks, and it has nothing to do with how many *paid* rest breaks class members took.

159. Dr. Phillips' deposition testimony-based unpaid rest break analyses are also unreliable because the Plaintiffs' recollections may be unreliable. As discussed above, there is no reason to

assume that class members' recollections of the frequencies and durations of their unpaid rest breaks are accurate or that survey respondents were representative of the class. Interestingly, Dr. Phillips' estimates are 210% to 450% bigger when based on "frequencies per day" than when they are based on "frequencies per 10 usual trips." Such a huge disparity calls to mind the Federal Reserve Bank of Boston results discussed above where respondents' estimated frequencies varied by a huge amount based on how the question was posed. The fact that deposition testimony regarding the frequency of rest breaks (and the damages estimates that rely upon them) are so sensitive to how the question is framed in terms of breaks per unit of time or breaks per number of trips suggests that any estimate from these deponents is unreliable. There is no reliable basis to determine whether it is the time-based deponents, the trip-based deponents or both whose answers are grossly dissimilar from the class average experience, but one of these three is necessarily the case.

160. Dr. Phillips' damages estimates regarding unpaid rest breaks are further unreliable because they ignore that Wal-Mart's rest break obligations are limited to ten minutes per four hours of work or major fraction thereof. Essentially, Dr. Phillips' analysis penalizes Wal-Mart for allowing drivers to take more than 10 minutes of rest for every 4 hours of work. The survey asked respondents about number of times drivers stopped driving specifically to take an unpaid rest break per 10 usual trips, and defined a trip as a journey that begins and ends at a distribution center. To my knowledge there is no reliable method given the data in the discovery record to determine what the deposition testimony or survey answers mean in terms of unpaid rest breaks taken per four hours worked. Trips may be under 100 miles or over 2,500 miles. A class member may have several trips a day or just one or two trips per week. Under these circumstances, it is impossible to convert estimates of rest time taken per 10 usual trips into estimates of rest time taken every four hours. Consequently, deposition testimony and survey responses concerning frequency and duration of unpaid rest breaks may correlate to more than 10 minutes of paid rest for every four hours of work. For example, the maximum response at deposition concerning frequency of rest breaks was two per usual trip, and the maximum response regarding usual break time duration was 30 minutes. Responses like these could easily correspond to more than 10 minutes of rest time per four hours of work depending on how long these particular respondents' "usual" trips were. In order to use the deposition testimony or survey responses in a way such that his estimates were not influenced by respondents who typically took more than 10



minutes of rest time per four hours of work, Dr. Phillips would have had to have gathered more information from the respondents about how long their trips were. He would also have had to conduct additional analyses to exclude from his analysis rest time in excess of 10 minutes per four hours of work.

161. To determine the minimum wage that was allegedly not paid for rest breaks, it would be necessary to know what days were lacking paid rest in order to determine the prevailing minimum wage. Of course, it would also be necessary to know how much paid rest was missing. As I discuss above, I am unaware of any method to do that on a class-wide basis.

162. Dr. Phillips' fifth rest break estimate was twenty minutes of pay at the prevailing minimum wage for each day that he determined that a class member worked. This estimate is inconsistent with the class member testimony that they took paid rest breaks so there was not necessarily a rest break violation on every workday.

#### **E. REFUELING AT WAL-MART LOCATIONS**

163. Dr. Phillips generated 12 estimates of the time that class members spent while their vehicles were being fueled at Wal-Mart locations. Four estimates were based on "GasBoy" data showing the frequency at which class members refueled at Wal-Mart locations in October 2014. One calculation was based on the average refuelings per trip according to GasBoy and the average refueling time from certain class members' deposition testimony; another was based on the median refueling frequency from GasBoy and the median refueling time from those class members' deposition testimony, the third was based on an arbitrary assumption that refueling took 15 minutes per occurrence and the fourth scaled the 15-minute estimate up by 8.76% purportedly to adjust for class members missing from the dispatch data.<sup>80</sup>

164. Dr. Phillips' GasBoy estimates are unreliable because there has been no testimony in this case about the reliability or accuracy of the specific data or what it represents, and there is no basis to extrapolate from one month of GasBoy data to an eleven year damages period. They are also unreliable for all the reasons discussed above in the context of pre-trip inspections, post-trip

---

<sup>80</sup> Dr. Phillips appears to have used the incorrect ratio to adjust for class members missing from the dispatch data. As characterized on pp. 7-9 of Dr. Phillips' report, 5.35% is the "ratio of layovers" that purportedly adjusts for class members missing from the dispatch data. 8.76% is the ratio of layovers found in the paper data to those found in the electronic dispatch data for those class members who appear in both data sets.



inspections and rest breaks regarding extrapolations from selected class members' deposition testimony to the class as a whole. These reasons concern recall bias, demand artifacts and absence of statistical validity. There is no basis to assume that Plaintiffs' testimony reliably estimates the average duration of past events. Of course, arbitrarily assuming 15 minutes is no better.

165. Even if the deposition testimony measured the typical duration of refueling accurately, it indicates gross disparities in refueling times. One respondent reported that it took him as little as 5 minutes while another respondent testified that it took him as much as 60 minutes, *i.e.*, 1100% more time.<sup>81</sup> Basing refueling damages estimates on the average or median from a few non-randomly selected class members or just a flat 15 minutes per person per event ignores the vast differences in survey respondents' reported refueling times. Finally, the GasBoy estimates are unreliable because they ignore class member testimony that they left their trucks and took paid or unpaid breaks while the trucks were being fueled as well as testimony that the trucks were fueled by people other than the drivers.<sup>82</sup>

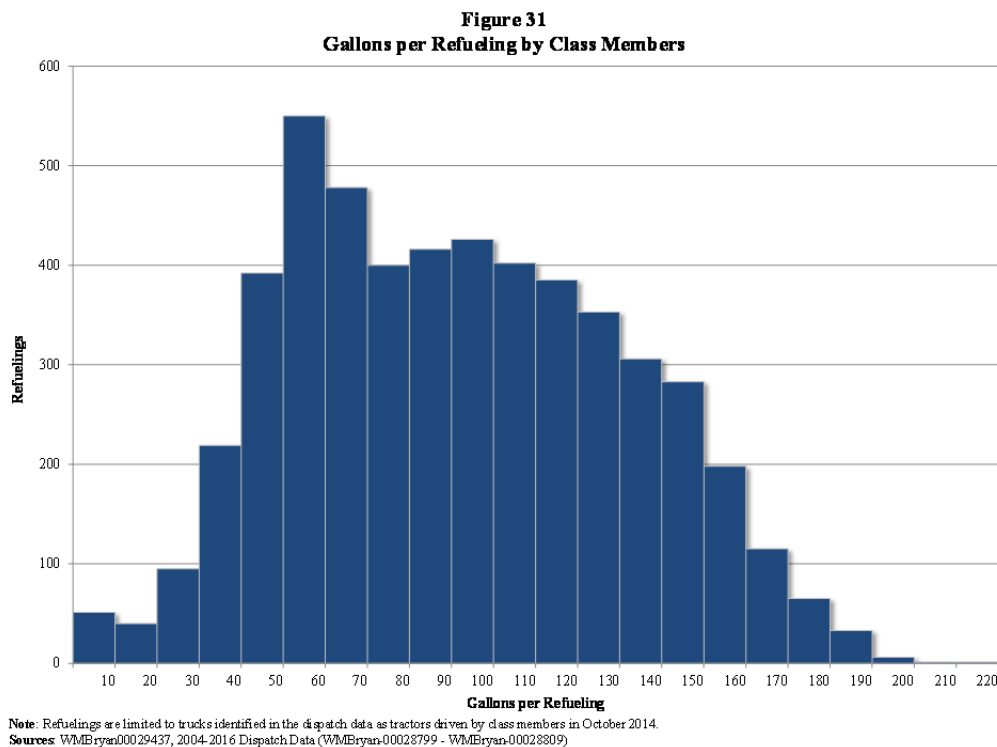
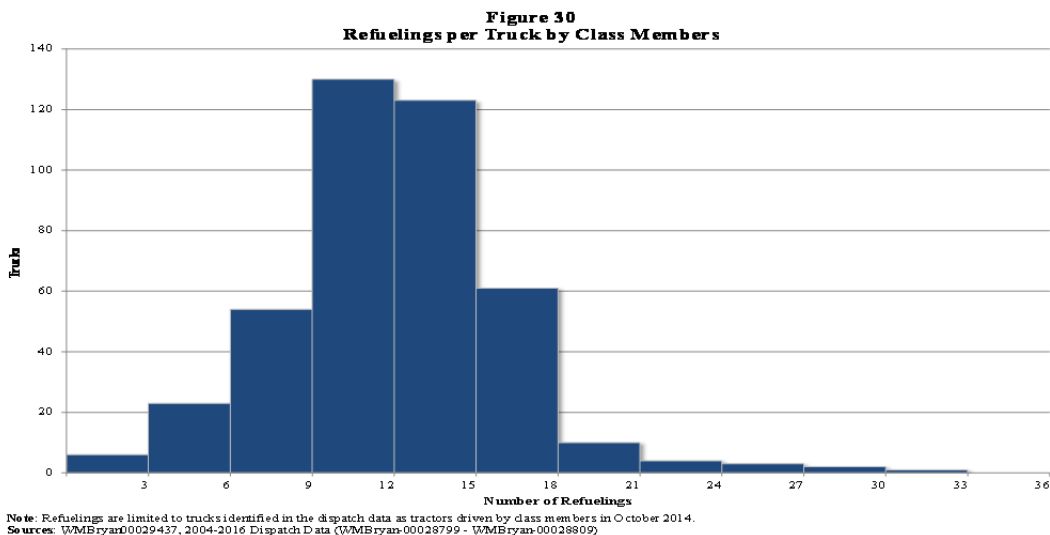
166. Dr. Phillips based four of his Wal-Mart location refueling estimates on deposition testimony relating to both the frequency and duration of refueling at Wal-Mart facilities. As with his rest break analysis, Dr. Phillips generated alternatives that were alternatively based on means or medians and whether deponents' testimony was framed in terms of frequency per period of time or frequency per 10 usual trips. These estimates are unreliable because, as I explain above, there is no reason to assume that the deposition testimony reliably measures frequency or duration of activities, because the samples of deponents are not random so extrapolations based on their testimony are statistically invalid, because such estimates fail to account for the differences in class members' individual experiences and because such estimates fail to deduct refueling time that occurred during paid or unpaid breaks.

167. Dr. Phillips based four of his Wal-Mart location refueling estimates on the assumption that Wal-Mart trucks get seven miles per gallon on average and that drivers would refuel every time they reached 25% of capacity. Dr. Phillips does not provide any references or sources

<sup>81</sup> Deposition of Daniel Steele 02/18/16, pp. 132:22-133:4; Deposition of Glen Craft 02/16/16, p. 53:18-20.

<sup>82</sup> See for example the deposition transcript of Robert Garcia at pages 127-128 wherein Mr. Garcia explains that he may have refueled during unscheduled time but that he cannot recall specific instances because using unscheduled time in that way would be "second nature."

indicating that these assumptions are true, and the GasBoy data suggest otherwise. The GasBoy data show a wide variation in the frequency at which trucks were refueled during the sampled month and a wide variation in the gallons of fuel dispensed per refueling event. See Figures 30 and 31.



168. If Dr. Phillips' assumptions about seven miles per gallon fuel efficiency and refueling every time a tank reached a quarter full were true, they would imply that each truck would need

to be refueled at regular mileage intervals depending on the truck's fuel capacity. As trucks are sometimes shared, it would not follow that each class member must have refueled at such regular intervals as Dr. Phillips assumes for purposes of his refueling damages estimates. In addition, the assumptions would imply the rate at which the trucks were refueled somewhere, not necessarily at Wal-Mart facilities. As we know that trucks were fueled outside of Wal-Mart as well as at Wal-Mart, and Dr. Phillips estimates damages related to this other refueling separately, Dr. Phillips' damages estimates would necessarily overstate time spent refueling at Wal-Mart if his other underlying assumptions about frequency and duration were true.

169. The four refueling damages estimates that are based on the 7 MPG/ 25% assumptions are distinguishable from each by their assumed time per refueling event: deposition testimony mean, deposition testimony median, 1 minute and 15 minutes. All of these duration estimates are unreliable. The two estimates based on median or mean durations from Plaintiffs' deposition testimony are unreliable for the reasons already discussed concerning recall bias, demand artifacts, statistical invalidity, failure to account for variation in duration across class members and failure to account for concurrent activities. Any estimate based on an arbitrary fixed amount of time, such as 1 minute or 15 minutes, is baseless by definition.

#### **F. OTHER TRIP BASED COMPUTATIONS**

170. Dr. Phillips presents minimum, median, average and maximum damages estimates related to the remaining activities covered by Plaintiffs' survey. These other activities are waiting time for unloading, waiting time for loading, CHP/DOT/weigh station inspection, truck washing, weighing outside of Wal-Mart, adjustments outside of Wal-Mart, fueling outside of Wal-Mart, and meeting time with driver coordinators at the beginning and end of each workday or trip. Dr. Phillips does not explain in detail how he arrived at these minimum, median, average and maximum estimates. Based on my review of Dr. Phillips' workpapers, these estimates are based on certain class members' deposition testimony concerning the frequency and duration of these activities. As with his other analyses, Dr. Phillips generated alternatives that were alternatively based on means or medians and whether deponents explicitly framed their frequency responses per units of time or per 10 usual trips. He generated additional alternatives for the activities of washing (based on the assumption that drivers wash their trucks once per week) and meeting

with driver coordinators at the beginning of each trip (one meeting for every trip).<sup>83</sup> Dr. Phillips calculates the median and average of the four to six alternative estimates for each activity and includes the minimum, median, average and maximum estimates in his report.

171. All of the damages estimates related to these remaining survey-related activities are unreliable for the various reasons discussed above regarding the unreliability of Dr. Phillips' other deposition-testimony –based estimates. I reiterate a few of these issues briefly.

172. There is no conceptual basis to assume that the deposition testimony would accurately reflect the deponents' actual experiences. The survey literature indicates that responses to questions like these tend to be systematically biased, often off by hundreds of percent on average even when respondents have a vested interest in responding accurately, e.g., when it relates to their health. There is no basis to assume that recollections given at deposition are going to be any more accurate.

173. The testimony does not necessarily account for activities that may occur simultaneously as paid activities. Many of the activities at issue could occur during time while drivers were being paid. The waiting time questions in particular may include time during which drivers were accruing live load, live unload and "unscheduled time" pay.<sup>84</sup>

174. Relatedly, Wal-Mart witnesses and class members have testified that some of these activities are performed as components of broader sets of paid activities. For example, deponents have testified that hook and depart pay is intended to compensate for all of the activities associated with departing with cargo from a facility including the pre-trip inspection, the pre-trip meeting with the distribution center coordinator and a certain amount of paperwork.

#### **G. SMALL SAMPLES – DISPARATE EXPERIENCES**

175. All of Dr. Phillips' various damages estimates that rely on deposition testimony are based on very small samples. Dr. Phillips only relied on 30 survey-related depositions to begin with,

---

<sup>83</sup> See Dr. Phillips' workpaper "Preliminary Loss Calculations by Sub-Category.xlsx".

<sup>84</sup> Upon review of Phillips' workpapers ("Deposition Responses.xlsx"), it appears that he based his damages estimates for loading and unloading *excluding* deposition testimony that explicitly framed the estimates as frequencies or durations of live loads, live unloads or grocery runs. However, his estimate includes testimony that is ambiguous as to whether the loading and unloading situations were live load, live unload or a different situation at a store or distribution center. (See, e.g., Deposition of Glen Craft 02/16/16, pp. 123:12-125:1.)

and not all deponents were asked about the duration and frequency of each activity.<sup>85</sup> One of Dr. Phillips' samples was comprised of only two class members; none included more than 27. It is implicit in Plaintiffs' own survey plan calling for 40 survey-respondents that none of these much smaller samples would be sufficiently large to draw reliable inferences even if they were randomly selected. As the sample size of 40 was presumably based on the expectation that durations and frequencies were fairly uniform across the class, and the necessary sample size to achieve a given margin of error increases as variability in the population increases, a much larger sample size than 40 would be needed to achieve the margin of error that Plaintiffs were seeking when they designed their survey plan.

176. The fact that Dr. Phillips' extraordinarily small samples of deponents were not randomly selected makes them even less reliable for purposes of drawing inferences about the class. Moreover, for each of the activity-related questions, there were wide disparities among deponents regarding durations and frequencies. Even if the averages and medians were based on random samples that were sufficiently large to generate acceptably small margins of error, most or all of the damages estimates would be grossly inaccurate because most individuals' experiences are far from average. Class members' experiences were not uniform so basing damages off of averages or medians will lead to gross inaccuracies at the individual level.

177. In Figure 32, I report the sample sizes that Dr. Phillips relies upon for each of the duration and frequency topics at issue and the range of responses by deponents. Figure 32 shows that half of Dr. Phillips' sample sizes were 13 or less. Figure 32 also shows tremendous heterogeneity across deponents in their reported frequencies and durations of the activities at issue in this case. For example, one deponent said he took zero to two rest breaks per day and another said he took 10. One deponent said he usually spent as little as a minute at end of trip meetings with driver coordinators and another said he usually spent as much as three hours. One deponent said he usually spent five minutes when he had to make adjustments after weighing and another said he usually spent two hours. Dr. Phillips' deposition-based analysis of the frequency per week of waiting for unloading was based on just two deponents. One deponent said he had to wait for unloading approximately once or twice per year. The other deponent said he had to wait for unloading approximately twice per week.

---

<sup>85</sup> See Dr. Phillips' workpaper "Deposition Responses.xlsx."

178. For completeness, I also report in Figure 32 maximum and minimum survey responses as Dr. Phillips interpreted them. The survey data reported in Figure 32 indicates the same heterogeneity across the class based on Dr. Phillips' interpretation of the survey responses as we see based on Wal-Mart business records, my interpretation of the survey responses and Dr. Phillips' selected deposition passages. There are hundreds and thousands of percentage point differences in answers about frequencies and durations of the activities at issue between survey respondents.

**Figure 32**  
**Deposition Testimony and Survey Responses Tabulated by Dr. Phillips**

Question	Deposition Responses				Survey Responses			
	Number of Numeric Responses	Minumum	Maximum	Range as % of Minimum	Number of Numeric Responses	Minumum	Maximum	Range as % of Minimum
3a Usual pretrip length	27	5.00	20.0	300%	28	5.00	30	500%
4a Usual post-trip length	25	4.00	15	275%	27	5.00	30	500%
5a # roadside/weigh inspections in usual month	21	0.00	1	Not Defined	16	0.00	15	Not Defined
5b Usual inspection length	24	10.00	45	350%	22	3.00	45	1400%
6a # washings in 10 usual trips	4	0.00	6	Not Defined	23	1.00	10	900%
6a # washings per week	24	0.00	4	Not Defined				Not Defined
6b Usual washing length	25	0.00	150	Not Defined	27	6.00	150	2400%
7a # weighings outside Wal-Mart in 10 usual trips	4	0.00	5	Not Defined	19	0.00	6	Not Defined
7a # weighings outside Wal-Mart per month	12	0.03	12	43100%				Not Defined
7b Usual length of weighing outside Wal-Mart	20	5.00	60	1100%	27	5.00	60	1100%
8a # adjustments outside Wal-Mart in 10 usual trips	6	0.50	2	300%	21	0.00	5	Not Defined
8a # adjustments outside Wal-Mart per year	8	0.33	156	46700%				Not Defined
8b Usual time to adjust	14	5.00	120	2300%	24	5.00	120	2300%
9a # refuelings at Wal-Mart in 10 usual trips	9	2.00	10	400%	25	1.00	10	900%
9a # refuelings at Wal-Mart per week	14	1.00	6	500%				Not Defined
9b Usual time to refuel at Wal-Mart	22	5.00	60	1100%	27	2.00	45	2150%
10a # refuelings outside Wal-Mart in 10 usual trips	10	0.00	10	Not Defined	23	0.00	10	Not Defined
10a # refuelings outside Wal-Mart per month	10	0.00	8	Not Defined				Not Defined
10b Usual time to refuel outside Wal-Mart	18	10.00	60	500%	25	10.00	60	500%
11a # meetings at start of day or trip in 10 usual trips	21	2.00	10	400%	26	2.00	11	450%
11a # meetings at start of day or trip per week	4	1.40	21	1400%				Not Defined
11b Usual length of meeting at start of day or trip	21	0.50	90	17900%	26	1.00	90	8900%
12a # meetings at end of trip in 10 usual trips	15	2.00	10	400%	25	1.00	10	900%
12a # meetings at end of trip per week	4	0.75	21	2700%				Not Defined
12b Usual length of meeting at end of trip	18	1.00	180	17900%	25	1.00	90	8900%
13a # times waiting for loading in 10 usual trips	6	1.00	6	500%	23	0.00	10	Not Defined
13a # times waiting for loading per week	4	0.50	9	1700%				Not Defined
13b Usual time waiting for loading	8	15.00	360	2300%	24	10.00	240	2300%
14a # times waiting for unloading in 10 usual trips	7	0.00	10	Not Defined	21	0.00	10	Not Defined
14a # times waiting for unloading per week	2	0.02	2	10300%				Not Defined
14b Usual time waiting for unloading	7	20.00	90	350%	22	20.00	240	1100%
15a # rest breaks in 10 usual trips	13	0.00	40	Not Defined	25	1.00	50	4900%
15a # rest breaks per day	12	0.00	10	Not Defined				Not Defined
15b Usual rest break length	17	5.00	30	500%	25	5.00	30	500%

Source: Dr. Phillips' Work paper "Deposition Responses.xlsx"

**Notes:**

1. The maximum and minimum deposition responses reported in this figure are the extremes that Dr. Phillips relied upon when calculating mean and median frequencies and durations for activities at issue in this case.
2. Dr. Phillips excluded deposition testimony responses if they were explicitly characterized as "live load", "live unload", "grocery", or "at a vendor".
3. When a deposition testimony response was "up to" or "less than" a number, Dr. Phillips sometimes took that number as the response (7b for David Lopez), and sometimes he treated the testimony as nonresponsive (question 11b for Ivan Harris).
4. For deposition testimony responses that contained more than two numbers, Dr. Phillips sometimes calculated a mean based on the minimum and maximum (11b for Charles McLaughlin) and treated that mean as the response. Other times, Dr. Phillips calculated a mean using all the numbers mentioned (15a for Robert Benavidez) and treated that mean as the response.
5. Dr. Phillips made some errors when calculating deposition testimony mean for some drivers, for example 11a for Todd Brown, 11a for Kenn Nevarez, and 13a for John Rivero.

## H. AGGREGATE DAMAGES

179. Dr. Phillips sums his damages estimates related to layovers, rest breaks and the various activities discussed above. For each category of damages, he provides a minimum, median, average and maximum estimate, so he has a total based on minimums, a total based on medians, a total based on averages and a total based on his maximums. As discussed, none of the individual estimates are reliable, so the totals are unreliable too. Moreover, Dr. Phillips also

overlooks that these various categories are not additive.

180. Survey respondents testified that many of these activities occurred concurrently with each other. Consequently, leaving aside the various errors discussed already, it would be inaccurate to add the various estimates together, as Dr. Phillips does, without further analysis to eliminate double-counting.

## **I. PENALTIES**

181. Dr. Phillips calculated penalties pursuant to California Labor Code Section 203 (“waiting time penalties”) and California Labor Code Section 1197.7 (“underpayment penalties”). I understand that waiting time penalties may accrue when employers fail to pay all compensation due at the time that an employment relationship is terminated. I understand that the potential penalty may be as much as a day’s pay per day of delay up to a maximum of 30 days’ pay. I understand that underpayment penalties may accrue when employees are paid less than the minimum wage. I understand that the penalty may be as much as \$100 per pay period per employee for the first violation per employer/employee relationship and as much as \$250 per pay period for subsequent violations.

182. In order to calculate penalties, Dr. Phillips assumed that “each component of alleged underpayment . . . occur[ed] on each pay period with the exception of layovers for each individual within the class.”<sup>86</sup> Based on this assumption, Dr. Phillips assumed that all class members were underpaid for each pay period and that terminated employees were also underpaid for their final paychecks. Absent this assumption, Dr. Phillips does not report any method for determining on a class-wide basis which class members if any were underpaid during any particular pay period or which class members if any were still owed additional compensation after having gotten their final paychecks. In other words, absent the assumption that every class member was underpaid every pay period, Dr. Phillips does not discuss any method of calculating penalties for the class.

183. Although it is the predicate for his penalty analysis, Dr. Phillips does not demonstrate that all class members suffered each category of underpayment during each of his or her pay periods or even that all class members suffered some underpayment during each of his or her pay periods. For most of the components of alleged underpayment, Dr. Phillips has literally no basis

---

<sup>86</sup> Phillips Report Appendix C (Garcia Report), of p. 6.

to assume that the activities occurred every working day. Plaintiffs' survey did not ask whether class members always engaged in the various tasks. Rather it asked whether they usually did it, whether they usually did it in 10 usual trips or whether they did it in a usual month. Moreover, for each of the activities on Plaintiffs' survey except for pre-trip inspections and washing the truck, the two activities that would seem most likely impacted by social desirability bias, at least one deponent testified that he did not usually engage in the activity.

184. Regarding pre-trip inspections, Wal-Mart and several class members testified that depart/hook pay is intended to compensate for pre-trip inspections, so pre-trip inspections are not necessarily unpaid according to this testimony.<sup>87</sup> Even assuming that every class member always conducted an initial pre-trip inspection and that all of these inspections were unpaid, Dr. Phillips has no basis to assume that each pre-trip inspection exceeded whatever threshold exists for *de minimis* activities under the Labor Code. Dr. Phillips simply has no reliable evidence regarding the time class members actually spent conducting pre-trip inspections. Even if the Plaintiffs' survey responses and deposition testimony are taken at face value, they are consistent with a large percentage of the pre-trip inspections conducted by the class taking five minutes or less.

185. Dr. Phillips assumes that each class member forewent two paid rest breaks on each workday, and he may ultimately argue as a fallback that this would justify an assumption that Wal-Mart underpaid every class member every pay period. However, class members testified that they were authorized and permitted to take paid rest breaks and that they took them. Consequently, Dr. Phillips would have to conduct an individualized inquiry to determine class member by class member and pay period by pay period whether Wal-Mart authorized and permitted the requisite number of paid rest breaks.

186. Dr. Phillips' failure to demonstrate that Wal-Mart underpaid any particular paycheck renders all of his penalty analyses unreliable. The penalty analyses are further flawed for penalty-specific reasons discussed below.

### ***Waiting time***

187. Dr. Phillips claimed to have identified 712 class members in the electronic dispatch data

---

<sup>87</sup> Depositions of Steven Davis (03/22/16) pp. 163:17-164:2, Robert Mosley (03/08/16) pp. 116:11-21 and 174:20-176:1, Michael Baker (04/15/16) p. 53:6-13 and Anthony Fantasia (03/07/16) p. 43:9-15; and the Depositions of drivers Stanley McCulley (04/28/16) pp. 57:17-58:22 and Angel Vasquez (05/04/16) pp. 24:8-26:6.



Wal-Mart produced in discovery. Those data include the date field “paid\_dt.” Dr. Phillips assumed that if the last entry in the paid\_dt field within the dispatch data occurred prior to October 16, 2015, then it was the class member’s date of separation of employment from Wal-Mart. Based on this assumption, Dr. Phillips identified 310 class members whom he assumed to be terminated from Wal-Mart. Having identified purportedly terminated employees, Dr. Phillips calculated their total pay for the preceding 13 pay periods and calculated their average daily pay. As Wal-Mart Fleet Drivers may either work seven days per pay period or 10 days per pay period, Dr. Phillips calculated average daily pay over the preceding 13 pay periods alternatively assuming the class member drove seven or 10 days per pay period. Dr. Phillips proceeded to calculate penalties for these class members assuming that all of them were underpaid at the assumed termination date. Dr. Phillips adjusts this penalty up to account for the 90 class members whom he could not find in the dispatch data. Dr. Phillips assumed that the same fraction of these 90 class members ( $310/712$ ) were terminated from Wal-Mart, that they had the same average daily pay as the 310 assumedly terminated class members in the dispatch data and that they too were entitled to the maximum penalty of 30 times average daily pay.

188. Dr. Phillips’ assumption was false that a final date entry earlier than October 16, 2015 in the “paid\_dt” field meant necessarily that a class member no longer worked for Wal-Mart. For example, employees may have such dates in this field if they are on leaves of absence. Two of the randomly selected class members were leaves of absence that commenced prior to October 16, 2015, Kenneth Nevarez and John Rivero. Their last entries in the paid\_dt fields were December 14, 2013 and December 3, 2011 respectively.<sup>88</sup> Determination that any of the other 308 persons whom Dr. Phillips assumed to be terminated actually was terminated would require further inquiry.

189. Dr. Phillips assumption is baseless that ( $310/712$ ) of the 90 class members not shown in the dispatch data must necessarily have been terminated. To my knowledge the only basis to assume that the 712 class members who are in the dispatch data are similar in all material respects to the 90 who are outside of it would be if inclusion or exclusion from the dispatch data

---

<sup>88</sup> There may be other explanations for the final date in the “paid\_dt” field being earlier than October 16, 2015. For example, drivers may transfer out of California or to other Wal-Mart jobs, and Dr. Phillips has not established how such job changes would be accounted for in the dispatch data.

was determined by a statistically random process. This assumption is implausible. More likely, there is some reason why the 90 class members are missing from the dispatch data and this reason could imply systematic differences between those 90 class members and the other 712 class members. Consequently, Dr. Phillips has no basis to assume similar termination rates between the 712 class members in the dispatch data and the 90 class members outside of it. Similarly, Dr. Phillips has no basis to assume similar rates of pay for whomever from among the 90 class members who did terminate, if anyone at all, and the 310 class members who were represented in the dispatch data whom Dr. Phillips assumed to be terminated from Wal-Mart.

### *Underpayment*

190. Given Dr. Phillips' assumption that every paycheck was underpaid, the underpayment penalty pertaining to any given class member depends on his length of employment as a Private Fleet Driver in California. Dr. Phillips has no information concerning the length of employment for the 90 class members not included in the dispatch data. Consequently, he assumes that penalties per class member for these 90 persons are equal to the median penalty he calculated for the 712 class members included in the dispatch data.<sup>89</sup> This assumption is baseless. Moreover, the limited information we do have suggests that the 90 class members who are not in the dispatch data were with Wal-Mart as Private Fleet Drivers in California for less time on average than the rest of the class members. Two of the 90 class members excluded from the dispatch data were surveyed and deposed. Based on their deposition testimony, their average tenure as Private Fleet Drivers in California during the class period was 43 months as of the time of their depositions.<sup>90</sup> Based on information available through November 28, 2014, the average tenure for the other 37 surveyed class members (excluding the class member who terminated from Wal-Mart immediately after hire) during the class period exceeded 99 months.<sup>91</sup> This calculation

---

<sup>89</sup> Phillips Report, Appendix C (Garcia Report), p. 7.

<sup>90</sup> Deposition of Michael Barker, 04/11/2016 pp. 18, 20, 36; Deposition of David Lopez, 03/03/2016, pp. 16, 29 and WM-BRYAN00029436. Mr. Lopez was domiciled in California from 2009 to June 2014. Mr. Barker was domiciled in CA from January 2012 to September 2013. Since the month in which Mr. Lopez was first based in California was unstated, I assume it was January 2009.

<sup>91</sup> WM-BRYAN00029436. Comprehensive data on hire and termination dates appears to end in November 2014. Driver Robert Benavidez was based out of Texas from July 2004 to September 2012 (Benavidez Deposition Exhibit A p. 1 and Benavidez Deposition p. 17:2-5). This period is excluded from the average tenure calculation. I am not aware of other survey respondents included in the dispatch data who were domiciled outside of California during the class period.

assumes that the California tenure of any surveyed class member lacking a termination date (i.e., still employed as of November 28, 2014) ended on that date. As such, it underestimates these 37 class members' average tenure in California during the class period.

## **IX. CONCLUSION**

191. Plaintiffs allege that Wal-Mart's policy was not to pay the minimum wage for layover time during which class members were supposedly under Wal-Mart's control, not to pay minimum wage for ten minutes of paid rest for each four hours of work or major fraction thereof and not to compensate for time spent on various particular activities such as waiting to depart or completing paperwork. Assuming that Wal-Mart had such written policies, there are inadequate data to prove that any individual class member suffered any injury because of it or to measure damages in any reliable way.

192. Some drivers took layovers away from their cabs at their discretion without seeking approval from Wal-Mart, and some said that there were no restrictions placed on them and no control by Wal-Mart even when they were in the truck on layover, so not all of the class appears to have been injured by the assumed Wal-Mart layover policy. Virtually all of the surveyed class members testified to spending layover time engaged in personal activities away from their trucks, even on occasions when they returned to the truck to sleep. Consequently, a reliable estimate of layover-related damages for those class members who were injured would have to account for time away from the truck during layovers and there are no data reliably measuring that time over the damages period. If layover time spent in the truck without restrictions imposed by Wal-Mart is not compensable, a reliable estimate of layover damages would have to account for that time too.

193. Wal-Mart apparently did authorize and permit rest breaks if paid rest breaks are defined to include rest time during which Wal-Mart was paying class members by the minute or activity. Many class members testified to taking paid rest breaks and to having the discretion to take breaks whenever they chose. Consequently, proof of injury apparently requires individualized proof. Dr. Phillips has not attempted to ascertain how many class members regularly took paid rest breaks or how many of the class members' four-hour work-periods included ten minutes or more of paid rest. I am unaware of any reasonably accurate and reliable method for doing so on a class wide basis.

194. The allegedly uncompensated activities occurred concurrently with each other and with paid activities. Dr. Phillips did not conduct any analysis attempting to demonstrate that each class member had uncompensated time after accounting for concurrent activities. Instead, he merely assumed that every class member suffered from every alleged violation, except for allegedly undercompensated layovers, every pay period. Not only is an assumption not economic proof, this assumption is demonstrably false. It conflicts with class members' testimony and survey responses including the deposition excerpts that Dr. Phillips relies upon that some class members never engaged in certain of the tasks at issue and others did so infrequently, e.g., once per year.

195. Even if injury related to allegedly uncompensated activities is assumed, there is no reasonably reliable and accurate way to measure damages without individualized evidence. Dr. Phillips presents no evidence at all about how much time was spent on the allegedly uncompensated activities after netting out time during which the class members were being compensated for something else and correcting the estimates so as not to double count time spent concurrently on multiple purportedly uncompensated activities, e.g., doing paperwork while waiting. To my knowledge, there is no good measure of the typical frequency or duration of any of the allegedly uncompensated activities. There was no such evidence in the Wal-Mart records I have reviewed. Plaintiffs' survey does not generate reliable estimates of these statistics due to high margins of error. The selected deposition testimony that Dr. Phillips relied upon is even worse as estimates based on selected deposition passages would also have a high margin of error even if they were statistically valid in the first place. Moreover, there is apparently significant variation across class members in how frequently they undertook the activities at issue and how long they spent on them. Even if class averages were estimated accurately and reasonably precisely, individual damages estimates would be highly inaccurate due to the heterogeneous experiences of the class members.

196. Speaking generally, Dr. Phillips' methodology is unreliable. He based his analysis on selected deposition testimony from as few as two class members to no more than 27, depending on the particular damages category and method. There is no valid statistical basis for Dr. Phillips to extrapolate from these small, nonrandom samples of class members to the class as a whole. Although it was a random process that generated the list of persons to whom Plaintiffs issued subpoenas to come take a survey and sit for deposition, it was not a random process that

determined who showed up and it was not a random process that determined which questions each surveyed class member would be asked at deposition. Former employees and employees on extended leaves of absence were overrepresented among survey respondent/deponents. These employees tended to say that activities were more frequent and took more time than current employees said. Overrepresentation of former employees apparently biased survey responses, and deposition testimony, leading towards overstated damages.

197. Plaintiffs' counsel posed their deposition questions after having access to deponents' survey responses, after hearing their answers to Wal-Mart's counsel's questions, in some cases after having reviewed deponents' answers to an earlier survey and potentially after having conferred privately with deponents subject to attorney client privilege. It is not unreasonable to expect that Plaintiffs' counsel may have selected deposition questions intending to elicit deposition testimony that is helpful to their case. Not knowing that Dr. Phillips would base his analysis on deposition testimony about frequency and duration (rather than survey data as Dr. Roberts and Mr. Chasworth said Plaintiffs would do) and not having the opportunity to confer with deponents privately, Wal-Mart's counsel would not have had the incentive or opportunity to fully offset the impact of such a strategy on the body of deposition testimony for Dr. Phillips to choose from. This is not to cast aspersions against Plaintiffs' counsel, but rather to identify a significant way in which the samples that Dr. Phillips relies upon are influenced by nonrandom factors. It is utterly implausible that Plaintiffs' counsel's and Wal-Mart's counsel's decisions about what questions to ask different deponents were determined randomly.

198. Leaving aside that the selected deponents' experiences are not statistically representative of the class, their recollections are not likely to be accurate anyway. It is well documented in the survey literature that people's recollections of the frequency of past activities are subject to recall bias so that even the average response across hundreds or thousands of survey respondents may be off by hundreds of percent. Similarly, marketing and other literature show that peoples' perceptions are poor measures of the duration of events. Validation studies indicate that only a small percentage of people measure the passage of time accurately without a clock. People tend to perceive wait time in particular as taking longer than it actually does. In the case at hand, the deponents admitted that their guesses were unreliable. There is no statistical basis to assume that the average guess is any more reliable than the individual guesses that feed into it.

199. The survey literature also warns about social desirability bias, which may reasonably influence how often people report performing safety inspections and washing trucks or how long they say they spend on these activities. The literature also speaks about demand artifacts and the difficulty, or impossibility, of a respondent not being influenced in his answers by knowledge about the reasons that the question is being asked. Here, it is implausible that none of the deponents whom Dr. Phillips relied upon had any knowledge that their answers might impact their future working conditions, their own finances, fellow drivers' finances and Wal-Mart's finances. The survey literature indicates that it is unreasonable to expect that these factors would have no influence on deponents' answers. That is why valid scientific research involving human subjects is generally conducted on a double-blind basis when possible.

200. Even if one were to ignore that it is statistically invalid to extrapolate the selected deponents' testimony to the class, that the deponents' testimony is unlikely to reflect accurately their own experiences and that deponents disavowed the reliability of their own estimates, there is too much variation in class members' experiences to draw reliable inferences based on sample sizes containing as few as two and no more than 27 members. Any estimate of the most typical frequency or most typical duration based on such a small sample would be subject to significant measurement error. Even if the most typical frequency or duration were measured accurately, individual damages would be inaccurate because of the heterogeneity across class members in their own durations and frequencies. In the case at hand, most typical is still not typical at all.

201. Dr. Phillips' penalty analysis is also unreliable. Both the waiting time penalty analysis and the underpayment penalty analysis are based on the assumption that each class member suffered from each of the alleged violations, except for underpaid layovers, every pay period. Dr. Phillips did not conduct any analysis to establish that this assumption is true, and it conflicts with class member testimony that there were some activities that they never did at all and others that they performed only infrequently. Dr. Phillips does not establish the number of pay periods, if any, for which any class member was underpaid. Similarly, Dr. Phillips does not establish that the final paycheck for any terminated employee was inaccurate. Dr. Phillips' penalty analyses are further flawed because they are based on invalid assumptions about the number of pay periods that class members worked and the number of class members who terminated from Wal-Mart during the class period.

202. Dr. Phillips states in his report that he anticipates conducting a significant amount of additional work, generating many new damages estimates and that he reserves the right to update or revise his analysis.<sup>92</sup> I understand that Wal-Mart may ask me to review and respond to any new analyses by Dr. Phillips if they are forthcoming and they are allowed by the Court. I have agreed to do so.

A handwritten signature in black ink that reads "Jonathan Walker". The signature is written in a cursive, flowing style. The first name "Jonathan" is written in a larger, more prominent script, and "Walker" is written in a slightly smaller, more compact script. The signature is positioned above a horizontal line.

Jonathan Walker

June 3, 2016

---

<sup>92</sup> Phillips Report, p. 3, 4, 11, 23.